

# **Subook**

**TO WHICH IS ADDED MANY USEFUL ONE  
LINERS AND CODE SO THAT  
YOU CAN AWK LIKE A HAWK BY**

**Changxing Su**

**An Human**

**DOHA  
PUBLISHED IN THE WILD**

# Perface

I am of the opinion that every  $\text{\LaTeX}$  geek, at least once during his life, feels the need to create his or her own class: this is what happened to me and here is the result, which, however, should be seen as a work still in progress. Actually, this class is not completely original, but it is a blend of all the best ideas that I have found in a number of guides, tutorials, blogs and [tex.stackexchange.com](http://tex.stackexchange.com) posts. In particular, the main ideas come from four sources:

- ▶ The [Kaobook](#), which was a model for the style.
- ▶ The [ElegantBook](#), which provided elegant and colorful theorem environment.
- ▶ *Molecular Biology of the Cell*, which provides a very good typesetting style for biology textbooks, is one of the reference objects of this template.
- ▶ [NotesTeX](#), which makes good use of the `marginnote` environment.

The first chapter of this book is introductory and covers the most essential features of the class. Next, there is a bunch of chapters devoted to all the commands and environments that you may use in writing a book; in particular, it will be explained how to add notes, figures and tables, and references. The second part deals with the page layout and design, as well as additional features like coloured boxes and theorem environments. Subsequent chapters provide writing examples for different subject contents.

I started writing this class as an experiment, and as such it should be regarded. Since it has always been intended for my personal use, it may not be perfect but I find it quite satisfactory for the use I want to make of it. I share this work in the hope that someone might find here the inspiration for writing his or her own class.

*Changxing Su*

# Contents

|                        |          |
|------------------------|----------|
| <b>Preface</b>         | <b>1</b> |
| <b>Contents</b>        | <b>2</b> |
| <b>List of Figures</b> | <b>4</b> |
| <b>List of Tables</b>  | <b>6</b> |

## I User Manual

|   |           |
|---|-----------|
| <b>1 Introduction</b>   | <b>8</b>  |
| 1.1 Required Packages   | 8         |
| 1.2 License   | 8         |
| 1.3 Features  | 8         |
| 1.4 $\text{\TeX}$ Shortcuts   | 9         |
| 1.5 <code>amsthm</code> Environments                                | 9         |
| 1.5.1 <code>tcolorbox</code> Environment and Known Issues . . . . . | 10        |
| 1.6 Fullpage Environment  | 10        |
| <b>2 Margin Stuff</b>   | <b>13</b> |
| 2.1 Sidenotes   | 13        |
| 2.2 Marginnotes   | 13        |
| 2.3 Why use both marginnotes and sidenotes?                         | 14        |

## II Biology notes

|  |           |
|--|-----------|
| <b>3 DNA, Chromosomes, and Genomes</b>                     | <b>16</b> |
| 3.1 The Structure of DNA Provides a mechanism for Heredity | 16        |

## III Machine Learning Notes

|  |           |
|--|-----------|
| <b>4 Machine Learning</b>              | <b>19</b> |
| 4.1 Review and Overview                | 19        |
| 4.2 Formulation of supervised learning | 19        |

|            |  |           |
|------------|--|-----------|
| <b>4.3</b> | <b>Regression problems and squared loss</b>          | <b>20</b> |
| <b>4.4</b> | <b>Linear regression under squared loss</b>          | <b>20</b> |
| <b>4.5</b> | <b>Parameterized families of hypotheses</b>          | <b>22</b> |
| 4.5.1      | Well-specified case and maximum likelihood . . . . . | 22        |
| <b>4.6</b> | <b>Training loss</b>                                 | <b>22</b> |
|            | <b>Bibliography</b>                                  | <b>24</b> |
| <b>A</b>   | <b>Fonts Testing</b>                                 | <b>25</b> |
| <b>A.1</b> | <b>Font Sizes</b>                                    | <b>25</b> |
| <b>A.2</b> | <b>Font Families</b>                                 | <b>25</b> |
| <b>B</b>   | <b>Heading on level 0 (chapter)</b>                  | <b>26</b> |
| <b>B.1</b> | <b>Heading on level 1 (section)</b>                  | <b>26</b> |
| B.1.1      | Heading on level 2 (subsection) . . . . .            | 26        |
| <b>B.2</b> | <b>Lists</b>   | <b>27</b> |
| B.2.1      | Example for list (itemize) . . . . .                 | 27        |
| B.2.2      | Example for list (enumerate) . . . . .               | 27        |
| B.2.3      | Example for list (description) . . . . .             | 28        |

# List of Figures

|     |  |    |
|-----|--|----|
| 1-1 | marginfigure .....   | 10 |
| 1-2 | Figure caption .....   | 10 |
| 1-3 | Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. ....  | 11 |
| 1-4 | Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. ....  | 11 |
| 1-5 | Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. ....  | 12 |
| 1-6 | Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. ....  | 12 |
| 2-1 | Marginfigure: Tikz .....   | 14 |
| 3-1 | The DNA double helix. (A) A space-filling model of 1.5 turns of the DNA double helix. Each turn of DNA is made up of 10.4 nucleotide pairs, and the center-to-center distance between adjacent nucleotide pairs is 0.34 nm. The coiling of the two strands around each other creates two grooves in the double helix: the wider groove is called the major groove, and the smaller the minor groove, as indicated. (b) A short section of the double helix viewed from its side, showing four base pairs. The nucleotides are linked together covalently by phosphodiester bonds that join the 3'-hydroxyl (–OH) group of one sugar to the 5'-hydroxyl group of the next sugar. Thus, each polynucleotide strand has a chemical polarity; that is, its two ends are chemically different. The 5' end of the DNA polymer is by convention often illustrated carrying a phosphate group, while the 3' end is shown with a hydroxyl ..... | 16 |

3-2 DNA as a template for its own duplication. because the nucleotide A successfully pairs only with T, and g pairs with C, each strand of DNA can act as a template to specify the sequence of nucleotides in its complementary strand. In this way, double-helical DNA can be copied precisely, with each parental DNA helix producing two identical daughter DNA helices.

17

3-3 The relationship between genetic information carried in DNA and proteins. . . . . 17

# List of Tables

|     |                                  |    |
|-----|----------------------------------|----|
| 1-1 | Sophisticated margin table ..... | 11 |
| 2-1 | Margintable .....                | 14 |

# **Part I**

## **User Manual**



# Introduction

As the year went on, I started typesetting my personal notes during class and realized that the  $\LaTeX$  format, while great for publications and lecture notes in general, was lacking a few small but useful template for me.

## 1.1 Required Packages

For *Subook*, the following packages are required

marginnote, sidenotes, fancyhdr, titlesec, geometry, and tcolorbox.

For a brief summary, the marginnote, sidenote, titlesec, and tcolorbox packages are used in creating the `\part` environment, the package geometry is used globally to set the page width, page height, and margin width, and finally, fancyhdr, which is overridden on the title page, the contents page, and the `\part` page, sets the header for the body.

## 1.2 License

This work may be distributed and/or modified under the conditions of the LaTeX Project Public License, either version 1.3 of this license or (at your option) any later version. The latest version of this license is found in <http://www.latex-project.org/lppl.txt>, and version 1.3 or later is part of all distributions of LaTeX version 2005/12/01 or later. The current maintainer of this work is Changxing Su.

## 1.3 Features

*Subook* includes the following:

1. Several mathematics and physics packages.
2. Margins and margin environments for tables, figures, and asides.
3.  $\TeX$  shortcuts for various math scripts namely vector bold math, `mathbb`, `mathfrak`, and `mathcal`.
4. `amsthm` integrations and special environments for theorems, lemmas, proofs, definitions, examples, and remarks.
5. Stylized support for the `part` environment.
6. A `fullpage` environment that spans across the text width and the margin for longer equations and horizontal figures.

Each of these will be discussed in the following subsections.

## 1.4 T<sub>E</sub>X Shortcuts

*subook* comes built in with a minimal set of keyboard shortcuts for a few special characters. All of these shortcuts can be found in *subook.cls* just under

```
% -----  
%           User Created Commands  
% -----  
...
```

If one has their own macros then simply add it under this area.

Most people have their own shortcuts for commonly used mathematics, such as derivatives or integrals. For those looking for physics shortcuts, the excellent physics package (automatically included in *subook*) has possible everything that one can imagine.

## 1.5 amsthm Environments

amsthm environments are defined as usual being enclosed by `\begin{environment}` `\end{environment}` and most have been modified ostensibly from the original amsthm presets. Primarily, most environments, with the exception of the exercise environment, are now integrated with the wonderful tcolorbox package. Note that the counting for theorems and lemmas is distinct from the counting for definitions. Also note that the breakable for tcolorbox allows these environments to span multiple pages. All of these environment and the associated tcolorbox are provided by the code in *subook.cls* just under L<sup>E</sup>C<sup>T</sup>U<sup>R</sup>E.

```
% -----  
%           User Created Environments  
% -----  
...  
%% ----- tcolorbox -----  
...
```

Definition 1.1 (Test)

The definition environment

♣

Lemma 1.1 (Test)

The lemma environment

◇

Theorem 1.1 (Test)

The theorem environment

♡

Corollary 1.1 (Test)

The corollary environment

◇

Proposition 1.1 (Test)

The proposition environment

♠

Example 1

Test

The example environment

Proof.

The proof environment

□

Remark.

The remark environment

█



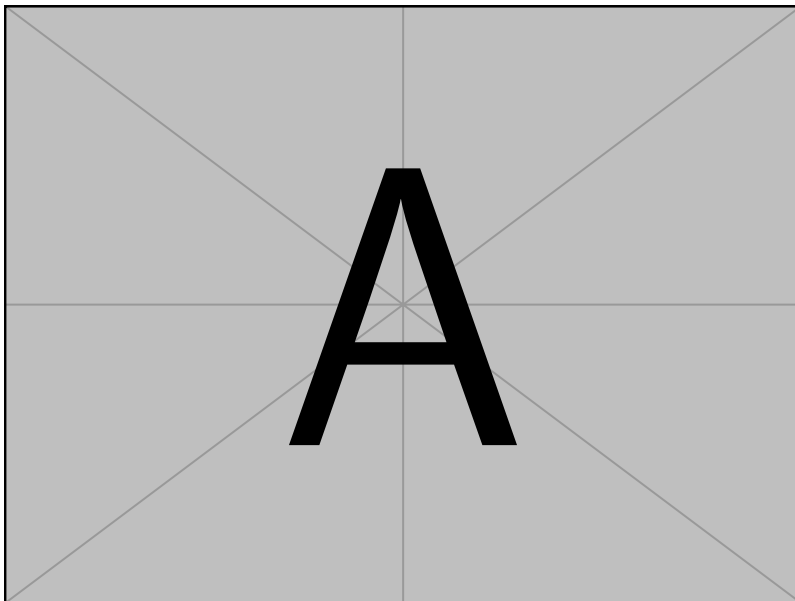
```
\captionof{figure}{Figure caption}
\label{fig:example} % Unique label used for referencing
the figure
```

The genomes of chimpanzees and humans are 99.9% identical, yet the differences between the two species are vast. The relatively few differences in genetic endowment must explain the possession of language by humans, the extraordinary athleticism of chimpanzees, and myriad other differences. Genomic comparison is allowing researchers to identify candidate genes linked to divergences in the developmental programs of humans and the other primates

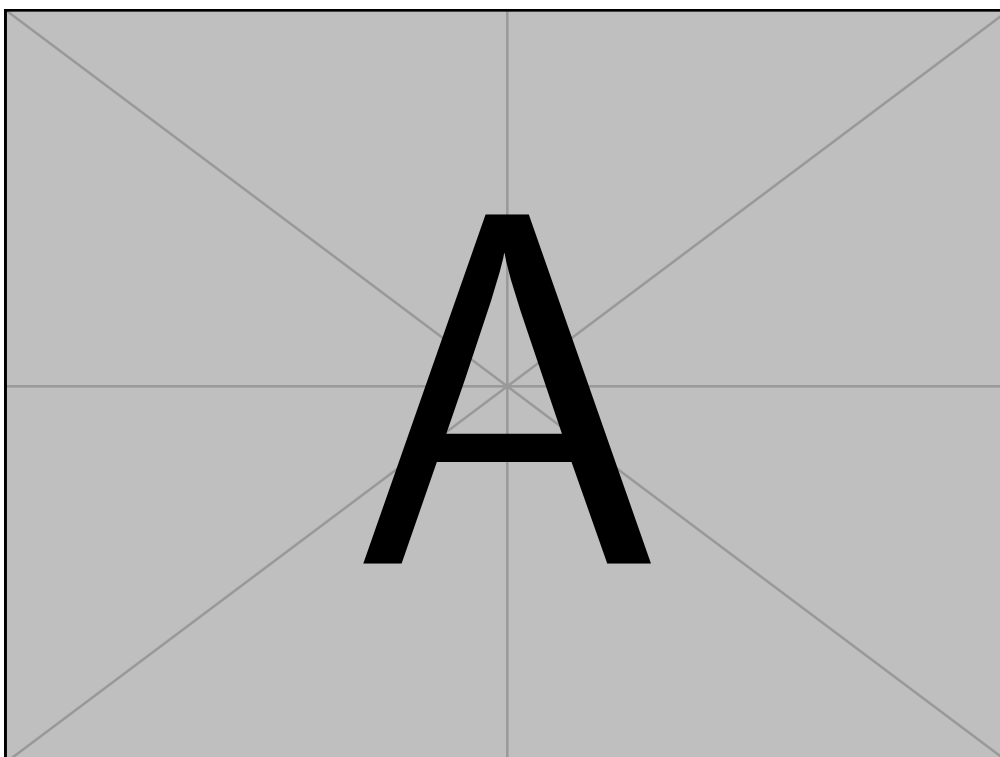
| 1 | 1  | 1     |
|---|----|-------|
| 1 | 1  | 0.562 |
| 2 | 1  | 0.910 |
| 3 | 11 | 0.296 |

**Table 1-1:** Sophisticated margin table

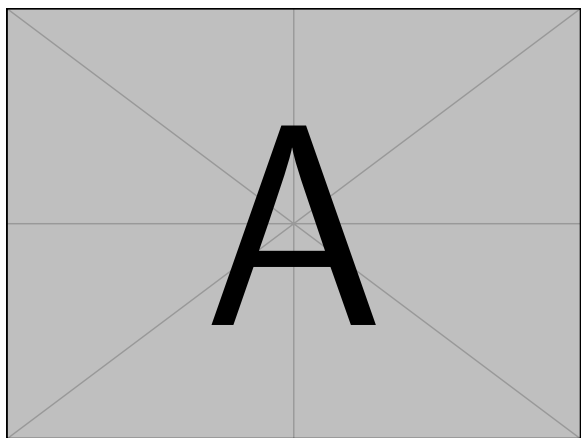
dsdsdsd



**Figure 1-3:** Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.

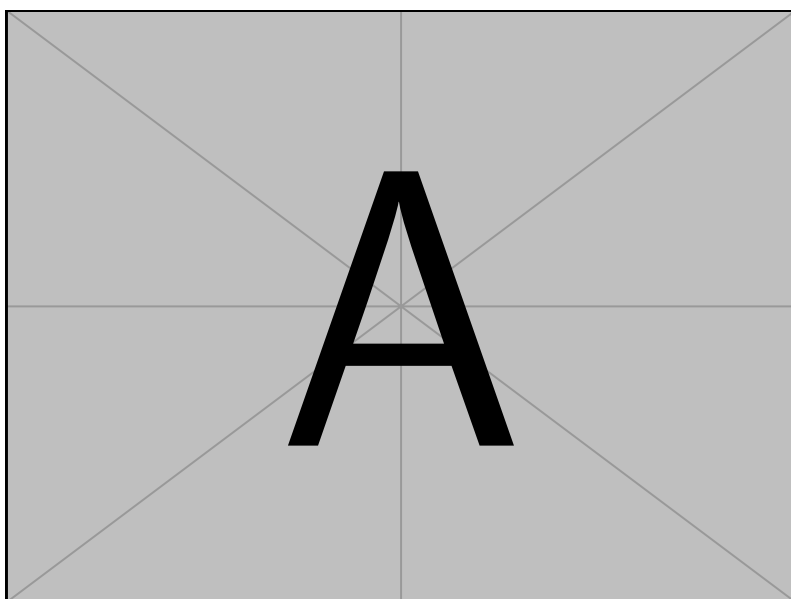


**Figure 1-4:** Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.



**Figure 1-6:** Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.

**Figure 1-5:** Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed. Camera mounted between two projections screens. Note that while the view direction can be modified, the up vector of the camera is fixed.



# Margin Stuff

Sidenotes are a distinctive feature of all 1.5-column-layout books. Indeed, having wide margins means that some material can be displayed there. We use margins for all kind of stuff: sidenotes, marginnotes, small tables of contents, citations, and, why not?, special boxes and environments.

## 2.1 Sidenotes

Sidenotes are like footnotes, except that they go in the margin, where they are more readable. To insert a sidenote, just use the command `sidenote{Text of the note}`. You can specify a mark<sup>o</sup> with `sidenote[mark]{Text}`, but you can also specify an offset, which moves the sidenote upwards or downwards, so that the full syntax is:

```
\sidenote [mark] [ offset ] { Text }
```

If you use an offset, you always have to add the brackets for the mark, but they can be empty.<sup>1</sup>

In subook we copied a feature from the `snotez` package: the possibility to specify a multiple of `baselineskip` as an offset. For example, if you want to enter a sidenote with the normal mark and move it upwards one line, type:

```
\sidenote [] [* -1] { Text of the sidenote. }
```

As we said, sidenotes are handled through the `sidenotes` package, which in turn relies on the `marginnote` package.

## 2.2 Marginnotes

This command is very similar to the previous one. You can create a marginnote with

```
marginnote[offset]{Text}
```

,where the offset argument can be left out, or it can be a multiple of

```
\marginnote [-12pt]{Text} or \marginnote [* -3]{Text}
```

Since `sidenotes` uses `marginnote`, what we said for marginnotes is also valid for sidenotes. Side- and margin- notes are shifted slightly upwards

```
(renewcommand{\marginnotevadjust}{3pt})
```

in order to align them to the bottom of the line of text where the note is issued. Importantly, both sidenotes and marginnotes are defined as floating if the optional argument ( the vertical offset) is left blank, but if the offset is specified they are not floating. Recall that floats cannot be nested, so in some rare cases you may encounter errors about lost floats; in those cases, remember that

<sup>o</sup> This sidenote has a special mark, a big O!

<sup>1</sup> If you want to know more about the usage of the `sidenote` command, read the documentation of the `sidenotes` package.

While the command for margin notes comes from the `marginnote` package, it has been redefined in order to change the position of the optional offset argument, which now precedes the text of the note, whereas in the original version it was at the end. We have also added the possibility to use a multiple of `baselineskip` as offset. These things were made only to make everything more consistent, so that you have to remember less things!

sidenotes and marginnotes are floats. To solve the problem, it may be possible to transform them into non-floating elements by specifying an offset of 0pt.

### 2.3 Why use both marginnotes and sidenotes?

Quite simply, marginnotes overlap each other if they are too close. This means that figures, and tables can overlap by just using marginnotes. This is why sidenotes is so useful as it not only numbers all side notes, but also dynamically aligns all side notes, figures, and tables.

So clearly, sidenotes must be better right? There are a few places where sidenotes fails too however. For instance, sidenotes cannot be used in equations, multicol, and with the tcolorbox for more details.environment. As the majority of the special environments from amsthm are modified to use tcolorbox, marginnotes becomes an essential part of NotesTeX.

The implementation of each of these is as follows.

- 1. Marginnote: This is how a `\marginnote{...}` behaves.
- 2. Mn: This is how a `\mn{...}` behaves.
- 3. Sidenote: This is how a `\sidenote{...}` behaves.<sup>2</sup>
- 4. Sn: This is how a `\sn{...}` behaves.<sup>3</sup>
- 5. Marginfigure: This environment requires the `\begin{marginfigure} ... \end{marginfigure}` enclosings. The caption package is needed to caption the figure.
- 6. Margintable: This environment requires the `\begin{margintable} ... \end{margintable}` enclosings. A table package, such as tabular, tabulary, tabu, or tabularx is required. The caption package is needed to caption the table.

Not numbered, 10pt.  
Numbered, footnotesize.  
<sup>2</sup> Numbered, 10pt.  
<sup>3</sup> Numbered, footnotesize.

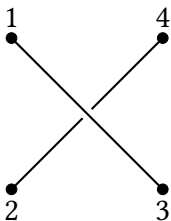


Figure 2-1: Marginfigure: Tikz

|          |        |
|----------|--------|
| NotesTeX | rocks! |
|----------|--------|

Table 2-1: Margintable

# **Part II**

## **Biology notes**



# DNA, Chromosomes, and Genomes

The documents of biology, chemistry and other disciplines are characterized by more legends and noun explanations, and less formulas. Here we have created an command sidebysidecaption to freely adjust the size of legends and figures. We use a portion of Chapter 4 of *Molecular Biology of the Cell*<sup>[1]</sup> for our template presentation.

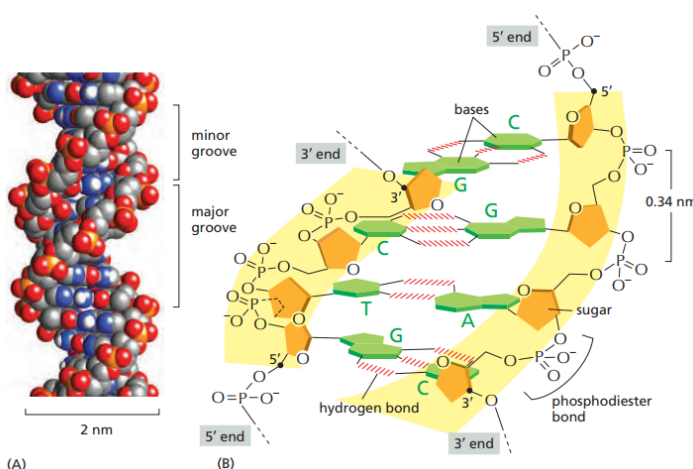
<sup>[1]</sup>Bruce Alberts *Molecular Biology of the Cell*

## 3.1 The Structure of DNA Provides a mechanism for Heredity

The discovery of the structure of DNA immediately suggested answers to the two most fundamental questions about heredity. First, how could the information to specify an organism be carried in a chemical form? And second, how could this information be duplicated and copied from generation to generation?

The answer to the first question came from the realization that DNA is a linear polymer of four different kinds of monomer, strung out in a defined sequence like the letters of a document written in an alphabetic script.

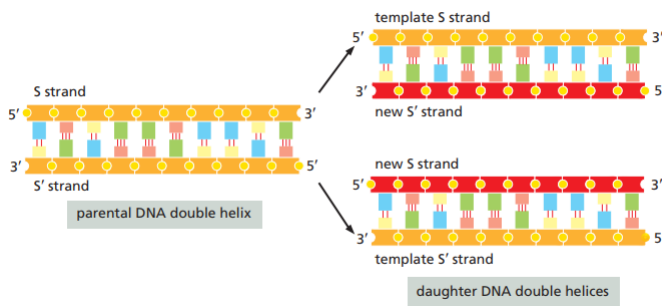
The answer to the second question came from the double-stranded nature of the structure: because each strand of DNA contains a sequence of nucleotides that is exactly complementary to the nucleotide sequence of its partner strand, each strand can act as a **template**, or mold, for the synthesis of a new complementary strand. In other words, if we designate the two DNA strands as S and S', strand



**Figure 3-1:** The DNA double helix. (A) A space-filling model of 1.5 turns of the DNA double helix. Each turn of DNA is made up of 10.4 nucleotide pairs, and the center-to-center distance between adjacent nucleotide pairs is 0.34 nm. The coiling of the two strands around each other creates two grooves in the double helix: the wider groove is called the major groove, and the smaller the minor groove, as indicated. (b) A short section of the double helix viewed from its side, showing four base pairs. The nucleotides are linked together covalently by phosphodiester bonds that join the 3'-hydroxyl (–OH) group of one sugar to the 5'-hydroxyl group of the next sugar. Thus, each polynucleotide strand has a chemical polarity; that is, its two ends are chemically different. The 5' end of the DNA polymer is by convention often illustrated carrying a phosphate group, while the 3' end is shown with a hydroxyl

S can serve as a template for making a new strand S', while strand S' can serve as a template for making a new strand S (Figure 3.1 on the facing page). Thus, the genetic information in DNA can be accurately copied by the beautifully simple process in which strand S separates from strand S', and each sepa-

rated strand then serves as a template for the production of a new complementary partner strand that is identical to its former partner.



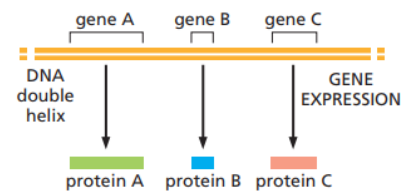
**Figure 3-2:** DNA as a template for its own duplication. because the nucleotide A successfully pairs only with T, and g pairs with C, each strand of DNA can act as a template to specify the sequence of nucleotides in its complementary strand. In this way, double-helical DNA can be copied precisely, with each parental DNA helix producing two identical daughter DNA helices.

The ability of each strand of a DNA molecule to act as a template for producing a complementary strand enables a cell to copy, or replicate, its genome before passing it on to its descendants. We shall describe the elegant machinery that the cell uses to perform this task in Chapter 5.

Organisms differ from one another because their respective DNA molecules have different nucleotide sequences and, consequently, carry different biological messages. But how is the nucleotide alphabet used to make messages, and what do they spell out?

As discussed above, it was known well before the structure of DNA was determined that genes contain the instructions for producing proteins. If genes are made of DNA, the DNA must therefore somehow encode proteins (Figure 3-3).

As discussed in Chapter 3, the properties of a protein, which are responsible for its biological function, are determined by its three-dimensional structure. This structure is determined in turn by the linear sequence of the amino acids of which it is composed. The linear sequence of nucleotides in a gene must therefore somehow spell out the linear sequence of amino acids in a protein. The exact correspondence between the four-letter nucleotide alphabet of DNA and the twenty-letter amino acid alphabet of proteins—the genetic code—is not at all obvious from the DNA structure, and it took over a decade after the discovery of the double helix before it was worked out. In Chapter 6, we will describe this code in detail in the course of elaborating the process of gene expression, through which a cell converts the nucleotide sequence of a gene first into the nucleotide sequence of an RNA molecule, and then into the amino acid sequence of a protein.



**Figure 3-3:** The relationship between genetic information carried in DNA and proteins.

# **Part III**

## **Machine Learning Notes**

# Machine Learning

In this chapter

Review and Overview  
Formulation of supervised learning  
Regression problems and squared loss  
Linear regression under squared loss

## 4.1 Review and Overview

In this lecture we delineate a mathematical framework for supervised learning. We focus on regression problems and define the notion of the loss/risk associated with a model. We then analyze a particular loss function (the squared loss) in a general setting, then specialize our result to the case of linear models. We next define the notion of parameterized families of hypotheses and the maximum likelihood estimate, and we conclude with an asymptotic result relating the training MLE to the true maximum likelihood parameter.

## 4.2 Formulation of supervised learning

We begin by constructing a mathematical framework for prediction problems. Our framework consists of the following elements:

1. A space of possible data points  $\mathcal{X}$ .
2. A space of possible labels  $\mathcal{Y}$ .
3. A joint probability distribution  $P$  on  $\mathcal{X} \times \mathcal{Y}$ . We assume that our training data consists of  $n$  points

$$(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)}) \stackrel{\text{i.i.d.}}{\sim} P$$

each drawn independently from  $P$ .

4. A prediction function/model  $f : \mathcal{X} \rightarrow \mathcal{Y}$ .
5. A loss function  $\ell : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ . We will usually assume that  $\ell$  is bounded below by some constant, typically 0.

Given the prediction function  $f$  and the loss function  $\ell$ , the loss of an example is  $\ell(f(x), y)$ . We can then define the *expected risk* (or *expected loss*, or *population risk*)

$$L(f) \triangleq \mathbb{E}_{(x,y) \sim P} [\ell(f(x), y)].$$

Our goal will be to obtain a small expected loss. Often it will be infeasible to consider all possible models  $f$ , so we may restrict ourselves to a certain family of hypotheses  $\mathcal{F}$ . In this case, we define the *excess risk* of a model  $f$  as

$$L(f) - \inf_{g \in \mathcal{F}} L(g).$$

This gives us a measure of how well our model fits the data relative to the best we can hope to do within our set of options  $\mathcal{F}$ .

Within this framework, there are two main types of problems we will consider: *regression* problems, where the set of labels is  $\mathcal{Y} = \mathbb{R}$ ; and *classification* problems, where the set of labels is some finite set  $\mathcal{Y} = \{1, \dots, k\}$ . We will focus on regression problems in this lecture.

## 4.3 Regression problems and squared loss

We consider the regression problem of predicting  $y$  given  $x$ . We take as our loss function the *squared loss*

$$\ell(\hat{y}, y) = (\hat{y} - y)^2, \quad L(f) = \mathbb{E}_{(x,y) \sim P}[(f(x) - y)^2].$$

In this setting, we can decompose the risk in a very informative way.

### Lemma 4.1 (Decomposition of loss)

Under the squared loss, we have the decomposition

$$L(f) = \mathbb{E}_{x \sim P_x}[(f(x) - \mathbb{E}[y | x])^2] + \mathbb{E}_{x \sim P_x}[\text{Var}(y | x)]$$

where  $P_x$  is the marginal distribution of  $x$ . ◇

The second term in this expansion is the intrinsic variable of the label; it gives a lower bound on the loss we can achieve. Since the first term in the decomposition is nonnegative, it is an immediate corollary that the optimal model is  $f(x) = \mathbb{E}[y | x]$ .

In order to prove Lemma 1, we make use of the following claim.

If  $Z$  is a random variable and  $a$  is a constant, then

$$\mathbb{E}[(Z - a)^2] = (\mathbb{E}[Z] - a)^2 + \text{Var}(Z).$$

The proof of this claim is left as an exercise on HW 0. We are now ready to prove Lemma 1.

**Proof.** [Proof of Lemma 4.1] We have

$$\begin{aligned} L(f) &= \mathbb{E}[(f(x) - y)^2] \\ &= \mathbb{E}_{x \sim P_x}[\mathbb{E}_{P_{y|x}}[(f(x) - y)^2 | x]] && \text{(Law of total expectation)} \\ &= \mathbb{E}_{x \sim P_x}[(f(x) - \mathbb{E}[y | x])^2 + \text{Var}(y | x)]. && \text{(Claim ??)} \end{aligned}$$

Note that Claim ?? holds in the third equation since  $f(x)$  is a constant when we have conditioned on  $x$ . The desired result follows from linearity of expectation. □

Lemma 4.1 gives us a general lower bound on risk under squared loss. If we impose more structure on the set of hypotheses  $\mathcal{F}$  from which we can select  $f$ , we can gain more information on the risk.

## 4.4 Linear regression under squared loss

A commonly used choice of hypotheses is the set of linear functions:

$$\mathcal{F} = \{f : \mathbb{R}^d \rightarrow \mathbb{R} \mid f(x) = w^\top x, w \in \mathbb{R}^d\}.$$

For  $f \in \mathcal{F}$ , we then have

$$L(f) = L(w) = \mathbb{E}[(w^\top x - y)^2].$$

Henceforth, we will denote  $w^* \in \operatorname{argmin}_{w \in \mathbb{R}^d} L(w)$  and  $\hat{w}$  will denote a model learned from training data.

One may ask why we have only allowed linear functions with 0 instead of allowing a nonzero intercept. Actually, the framework we have outlined above is enough to accomodate nonzero intercepts. If we wish to analyze the function  $w^\top x + b$ , we can simply set  $\tilde{x} = (x, 1)$  and  $\tilde{w} = (w, b)$ . Then  $\tilde{w}^\top \tilde{x} = w^\top x + b$  and we have reduced to the case of 0 intercept.

When we restrict to linear models, we can further decompose the risk under squared loss.

#### Lemma 4.2

With  $w^* \in \operatorname{argmin}_{w \in \mathbb{R}^d} L(w)$ , we have

$$L(\hat{w}) = \mathbb{E}_x[\operatorname{Var}(y|x)] + \mathbb{E}_x[(\mathbb{E}[y|x] - w^{*\top}x)^2] + \mathbb{E}_x[(w^{*\top}x - \hat{w}^\top x)^2]. \quad (4.1) \quad \diamond$$

The second term in equation (4.1) can be thought of as the approximation error incurred by linear models. The third term can be interpreted as the estimation error we incur from having only a finite sample.

**Proof.** Define  $g(\hat{w}) \triangleq \mathbb{E}[(\mathbb{E}[y|x] - \hat{w}^\top x)^2]$ . By Lemma 4.1,

$$L(\hat{w}) = \mathbb{E}[\operatorname{Var}(y|x)] + g(\hat{w}). \quad (4.2)$$

Observe that since  $w^* \in \operatorname{argmin} L(w)$ ,  $\nabla L(w^*) = 0$ . Furthermore, since  $\mathbb{E}_x[\operatorname{Var}(y|x)]$  is a constant with respect to  $w$ , we have

$$\begin{aligned} \nabla L(w) &= \nabla g(w) \\ &= \mathbb{E}[\nabla_w (\mathbb{E}(y|x) - w^\top x)^2] \\ &= 2\mathbb{E}[(\mathbb{E}(y|x) - w^\top x)x]. \end{aligned}$$

Since  $\nabla L(w^*) = 0$  we have

$$\mathbb{E}[(\mathbb{E}[y|x] - w^{*\top}x)x] = 0. \quad (4.3)$$

Next, we expand:

$$\begin{aligned} g(\hat{w}) &= \mathbb{E}[(\mathbb{E}[y|x] - \hat{w}^\top x)^2] \\ &= \mathbb{E}[(\mathbb{E}[y|x] - w^{*\top}x) - (\hat{w}^\top x - w^{*\top}x)]^2 \\ &= \mathbb{E}[(\mathbb{E}[y|x] - w^{*\top}x)^2 + (\hat{w}^\top x - w^{*\top}x)^2 \\ &\quad - 2\mathbb{E}[(\mathbb{E}[y|x] - w^{*\top}x)(\hat{w}^\top x - w^{*\top}x)]]. \end{aligned}$$

Finally, observe that

$$\mathbb{E}[(\mathbb{E}[y|x] - w^{*\top}x)(\hat{w}^\top x - w^{*\top}x)] = (\hat{w}^\top - w^{*\top})\mathbb{E}[(\mathbb{E}[y|x] - w^{*\top}x)x].$$

By equation (4.3), this quantity vanishes and it follows that

$$g(\hat{w}) = \mathbb{E}[(\mathbb{E}[y|x] - w^{*\top}x)^2] + \mathbb{E}[(\hat{w}^\top x - w^{*\top}x)^2]. \quad (4.4)$$

Combining equations (4.2) and (4.4) gives the desired result.  $\square$

## 4.5 Parameterized families of hypotheses

Linear models are one type of *parameterized family* of hypotheses. In general, a parameterized family is given by a parameter space  $\Theta$ . For each  $\theta \in \Theta$  there is a hypothesis  $f_\theta(x)$ , sometimes written  $f(\theta; x)$ . In this case we may write the loss function as

$$\ell(f_\theta(x), y) = \ell((x, y), \theta).$$

In the special case of linear functions, our parameter space is  $\Theta = \mathbb{R}^d$  and for  $\theta \in \Theta$  we have  $f_\theta(x) = \theta^\top x$ .

### 4.5.1 Well-specified case and maximum likelihood

In the well-specified case,  $P_\theta(y | x)$  is a family of distributions parameterized by  $\theta \in \Theta$ , and  $y | x \sim P_{\theta^*}(y | x)$  is distributed according to some ground truth parameter  $\theta^*$ . We define the *maximum likelihood* loss function by

$$\ell((x, y), \theta) = -\log P_\theta(y | x),$$

so that minimizing the loss function is equivalent to maximizing the likelihood of the data.

For example, suppose that  $y | x$  is Gaussian distributed with mean  $\theta^{*\top}x$  and variance 1, i.e.  $y | x \sim N(\theta^{*\top}x, 1)$ . The likelihood is then

$$\begin{aligned} \ell((x, y), \theta) &= -\log P_\theta(y | x) \\ &= -\log \exp\left(-\frac{(y - \theta^\top x)^2}{2}\right) + c \\ &= \frac{(y - \theta^\top x)^2}{2} + c \end{aligned}$$

where  $c$  is the log of the normalizing constant. This computation shows that in the Gaussian setting, minimizing the squared loss actually recovers the MLE.

## 4.6 Training loss

Often we do not know the true underlying distribution  $P$  with which to compute the expected loss. In these cases we need to use an approximation based on the data we do have. This motivates our definition of the *training loss*

$$\hat{L}(\theta) \triangleq \frac{1}{n} \sum_{i=1}^n \ell((x^{(i)}, y^{(i)}), \theta).$$

In the special case of maximum likelihood, we have  $\hat{L}(\theta) = -\frac{1}{n} \sum_{i=1}^n \log p_\theta(y^{(i)} | x^{(i)})$ . We define the *maximum likelihood estimator*

$$\hat{\theta}_{\text{MLE}} \in \operatorname{argmin}_{\theta \in \Theta} \hat{L}(\theta).$$

This approximation is “good” in the sense that as  $n \rightarrow \infty$ , the minimizer of the training loss  $\hat{\theta}_{\text{MLE}}$  approaches the true maximum likelihood parameter  $\theta^*$ . The following theorem quantifies this fact.


**Theorem 4.1 (Asymptotic of MLE)**

Assume  $\nabla^2 L(\theta^*)$  is full rank. Let  $\hat{\theta} = \hat{\theta}_{\text{MLE}}$  and

$$Q \triangleq \mathbb{E}_{(x,y) \sim P} [\nabla_{\theta}(\log p_{\theta}(y | x))(\theta^*) \nabla_{\theta}(\log p_{\theta}(y | x))(\theta^*)^{\top}].$$

Assuming that  $\hat{\theta} = \hat{\theta}_n \xrightarrow{p} \theta^*$  (i.e. consistency) and under appropriate regularity conditions,

$$\sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{d} N(0, Q^{-1}) \text{ and } n(L(\hat{\theta}) - L(\theta^*)) \xrightarrow{d} \frac{1}{2}\chi^2(p).$$

as  $n \rightarrow \infty$ , where  $p$  is the dimension of  $\theta$  and  $\chi^2(p)$  is the distribution of the sum of the squares of  $p$  i.i.d. standard Gaussian random variables. 

**Remark.** The positive definiteness of the Hessian  $\nabla^2 L(\theta^*)$  guarantees that identifiability holds *locally* (in a neighborhood of  $\theta^*$ ), but does not imply identifiability because of a lack of *global* information.

To give a counter-example, suppose  $\theta^*$  is the global minimizer of  $L$  and the Hessian is positive definite, but there exists a sequence of  $\theta_n$  such that  $\|\theta_n - \theta^*\| = n$  but  $L(\theta_n) = L(\theta^*) + 1/n$ , then the identifiability is violated: the inf is not strictly greater than but equal to  $L(\theta^*)$ . The reason is that the Hessian does not reveal information about  $L$  outside an infinitesimal neighborhood of  $\theta^*$ .

One way to exclude such adversarial case is to assume convexity: when  $L$  is convex, a local strong growth implies global growth and thus identifiability



# Bibliography

- [1] BRUCE ALBERTS A J LEWIS J. Molecular Biology of the Cell. [S.l.]: Garland Science, 2015.

# Fonts Testing

## A.1 Font Sizes

The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog.

## A.2 Font Families

The quick brown fox jumps over the lazy dog. Medium.

**The quick brown fox jumps over the lazy dog. Bold.**

The quick brown fox jumps over the lazy dog. Upright.

*The quick brown fox jumps over the lazy dog. Italics.*

*The quick brown fox jumps over the lazy dog. Slanted.*

THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG. SMALL CAPS.

The quick brown fox jumps over the lazy dog. Medium.

**The quick brown fox jumps over the lazy dog. Bold.**

The quick brown fox jumps over the lazy dog. Upright.

*The quick brown fox jumps over the lazy dog. Italics.*

*The quick brown fox jumps over the lazy dog. Slanted.*

THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG. SMALL CAPS.

The quick brown fox jumps over the lazy dog. Medium.

**The quick brown fox jumps over the lazy dog. Bold.**

The quick brown fox jumps over the lazy dog. Upright.

*The quick brown fox jumps over the lazy dog. Italics.*

*The quick brown fox jumps over the lazy dog. Slanted.*

THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG. SMALL CAPS.

## Heading on level 0 (chapter)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

### B.1 Heading on level 1 (section)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

#### B.1.1 Heading on level 2 (subsection)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

#### Heading on level 3 (subsubsection)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec

ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

**Heading on level 4 (paragraph)** Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

## B.2 Lists

### B.2.1 Example for list (itemize)

- ▶ First itemtext
- ▶ Second itemtext
- ▶ Last itemtext
- ▶ First itemtext
- ▶ Second itemtext

#### Example for list (4\*itemize)

- ▶ First itemtext
  - First itemtext
    - \* First itemtext
      - First itemtext
      - Second itemtext
    - \* Last itemtext
  - First itemtext
- ▶ Second itemtext

### B.2.2 Example for list (enumerate)

1. First itemtext
2. Second itemtext
3. Last itemtext
4. First itemtext
5. Second itemtext

### Example for list (4\*enumerate)

1. First itemtext
  - (a) First itemtext
    - i. First itemtext
      - A. First itemtext
      - B. Second itemtext
    - ii. Last itemtext
  - (b) First itemtext
2. Second itemtext

### B.2.3 Example for list (description)

**First** itemtext  
**Second** itemtext  
**Last** itemtext  
**First** itemtext  
**Second** itemtext

### Example for list (4\*description)

**First** itemtext  
    **First** itemtext  
        **First** itemtext  
        **Second** itemtext  
    **Last** itemtext  
    **First** itemtext  
**Second** itemtext