**DTS201TC Final Exam Review Guide**

# Lectures

***The list below hightlights important knowledge to support the final exam review.***

## Lecture 01

* + **what is pattern recognition**
  + Pattern recognition is the process of recognizing regularities in data
  + by a machine.
  + Pattern Recognition is the study of how machine can
  + • observe the environment
  + • learn to distinguish patterns
  + • make reasonable decisions about the categories of the patterns
  + **describe some real-life applications**

Speech RecognitionFinancial Trading

Object Detection/Classification

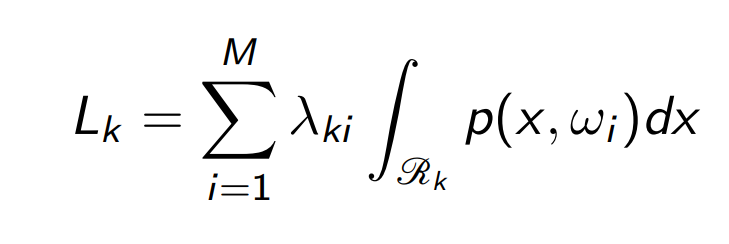
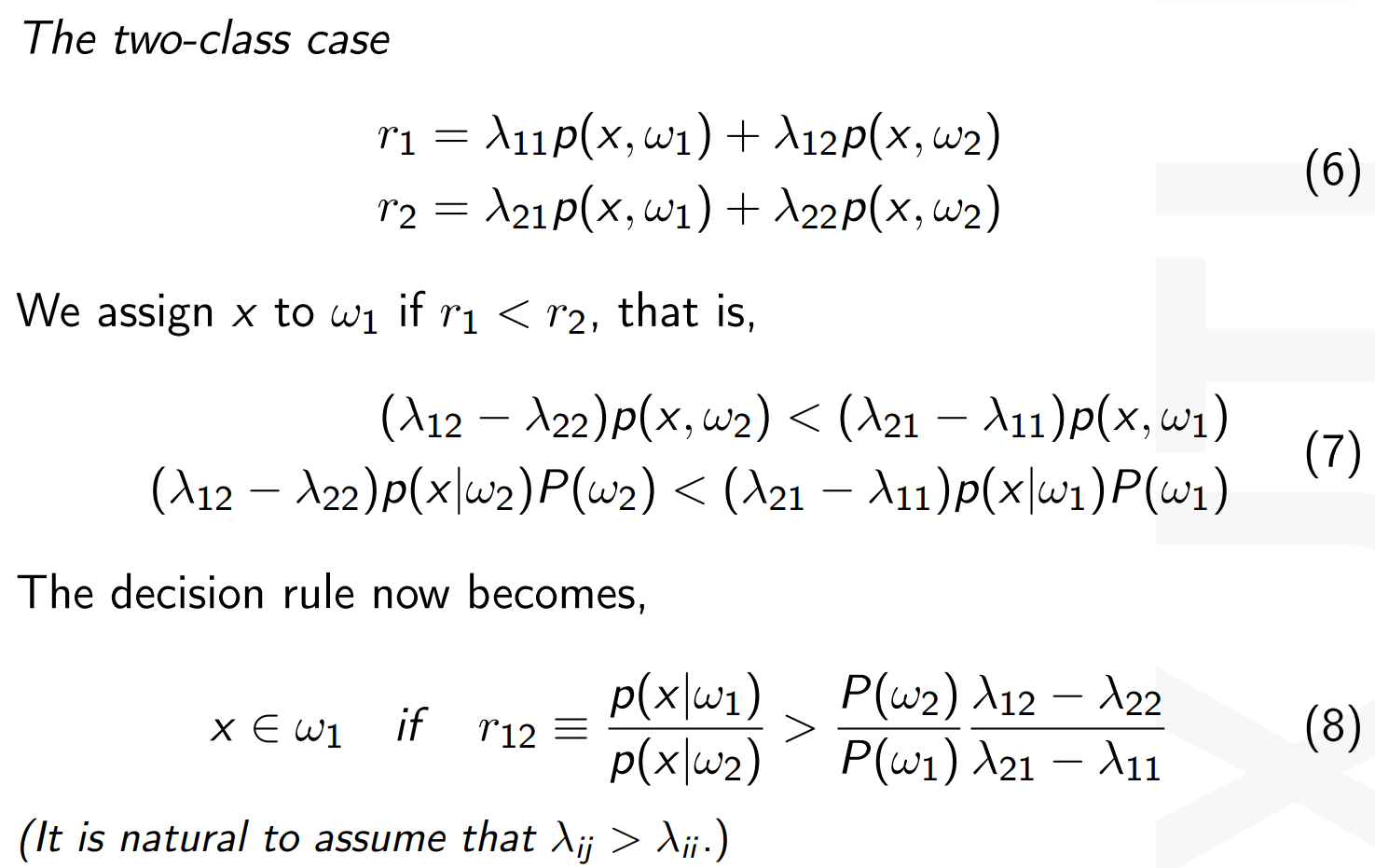
Fingerprint Identification

Face Recognition

Movie/Music Recommendation

* + **motivations (Fundamental Concepts, Ethical Considerations, Research and Innovation,etc.)**
  + Fundamental Concepts: Pattern Recognition provides the foundational understanding of how machines can learn from data and make decisions based on patterns. This knowledge is crucial for students who want to delve deeper into AI and machine learning. Understanding the underlying principles helps them grasp advanced techniques more effectively.
  + • Research and Innovation: Innovation often builds upon existing knowledge. While AI is rapidly advancing, new breakthroughs can still be rooted in Pattern Recognition principles. Students who understand these foundations are better positioned to contribute to research and push the boundaries of AI.
  + • Ethical Considerations: Pattern Recognition includes discussions on data bias, fairness, and privacy, which are critical aspects of responsible AI development. Ensuring that AI systems do not perpetuate biases and respect user privacy is essential, and a solid understanding of these concepts starts with Pattern Recognition.

## Lecture 04a

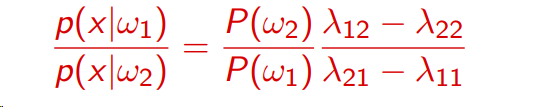
* + **Average Risk (Expected loss)**
  + **定义 公式**
  + 评估错误决定的成本度量
  + Average risk, or expected loss, is a measure used in pattern recognition to evaluate the cost or loss associated with incorrect decisions under uncertainty. It guides decision-makers in minimizing expected costs when choosing the best approach or model.  

**Exercises（注意输出为x0！）**

（1）最小错误概率

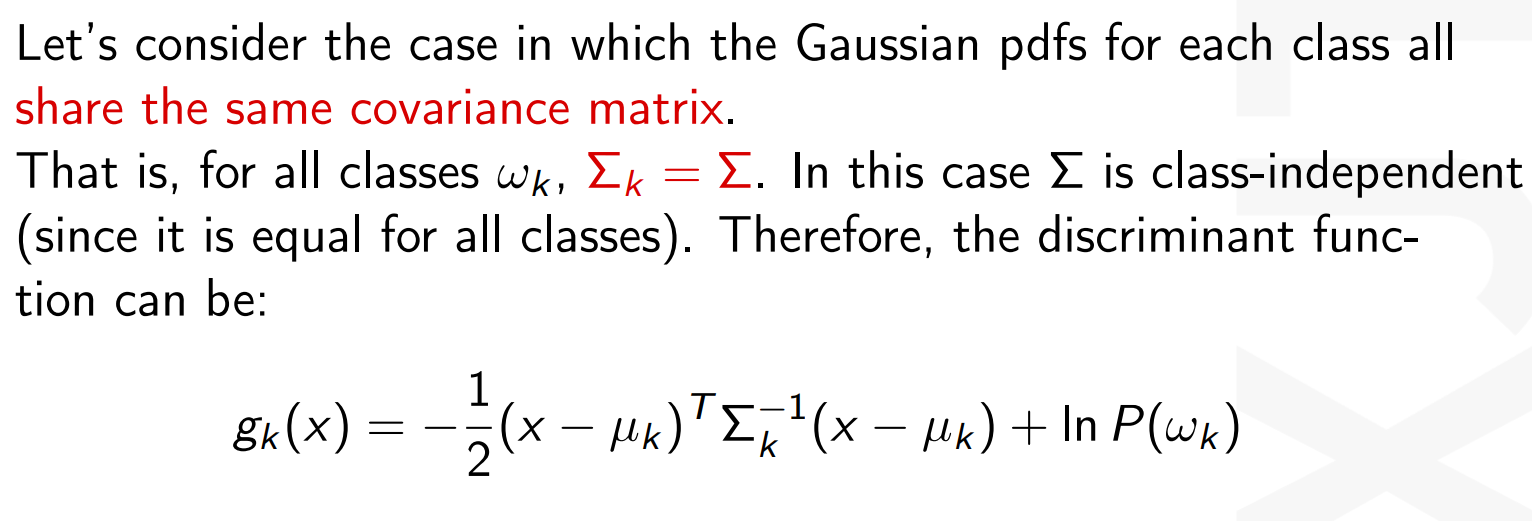


（2）最小风险



## Lecture05 under what conditions discriminant functions become linear

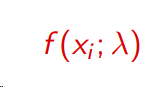
Discriminant 判别 covariance 协方差



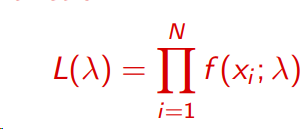
## Lecture 07 With data set given and *pdf* of the data known, how to calculate the *pdf* ’s parameters based on MLE approach.

Parameters 参数 MLE most likelihood extimation

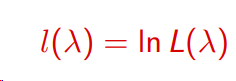
1. probability density fuction(罗列概率密度函数)

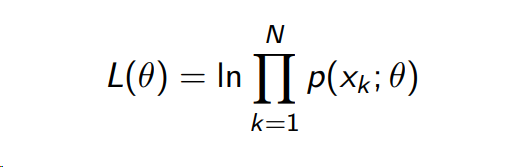


1. The likelihood function（写出似然函数）

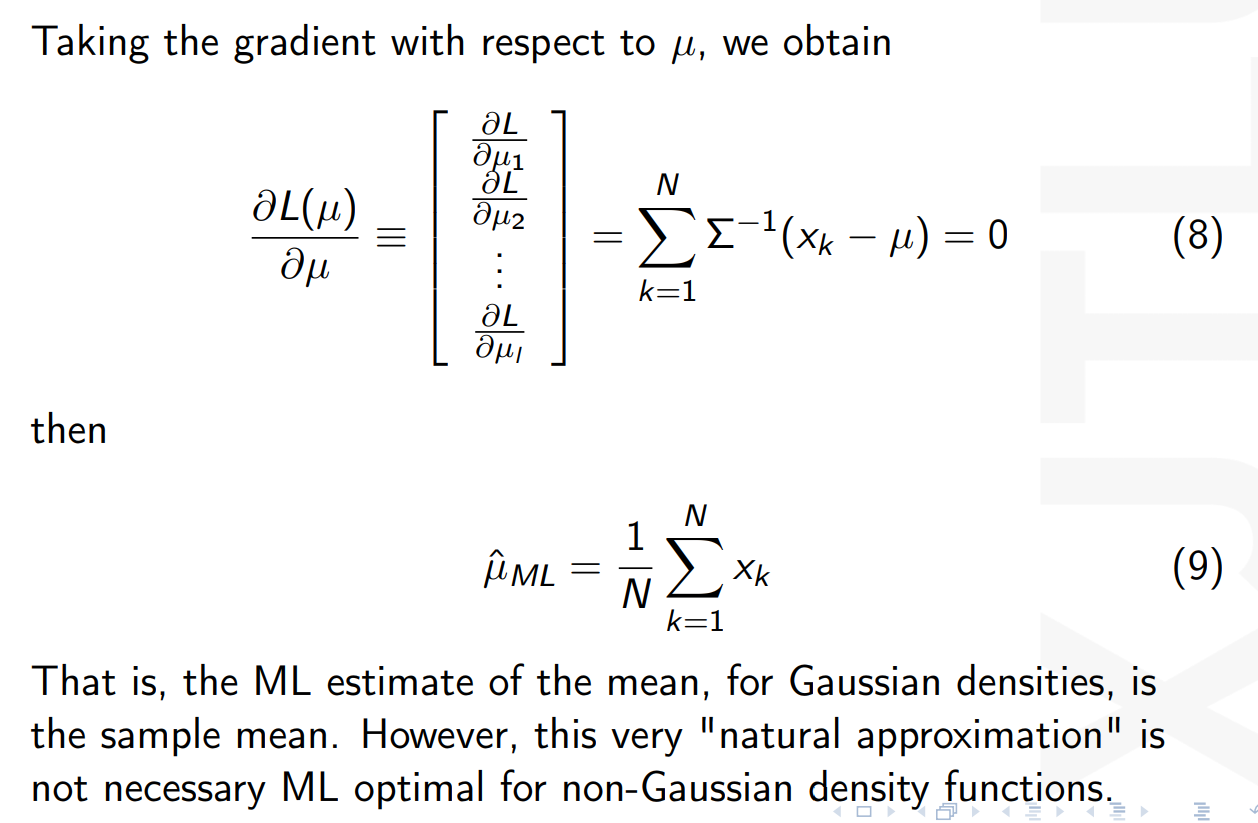


1. The natural log likelihood function（取对数）（有x项为Σ，无x项为N）

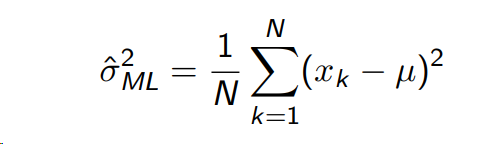


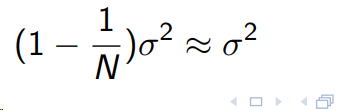


1. Derivate of The natural log likelihood function（对参数求导）
2. Set the derivative equal to zero（令导数等于0）
3. 字母 is a maximum extreme point（二阶导小于0）
4. 高斯分布中方差知均值不知

****

1. 高斯分布中均值知方差不知





3.二项分布的似然函数不需要累乘！



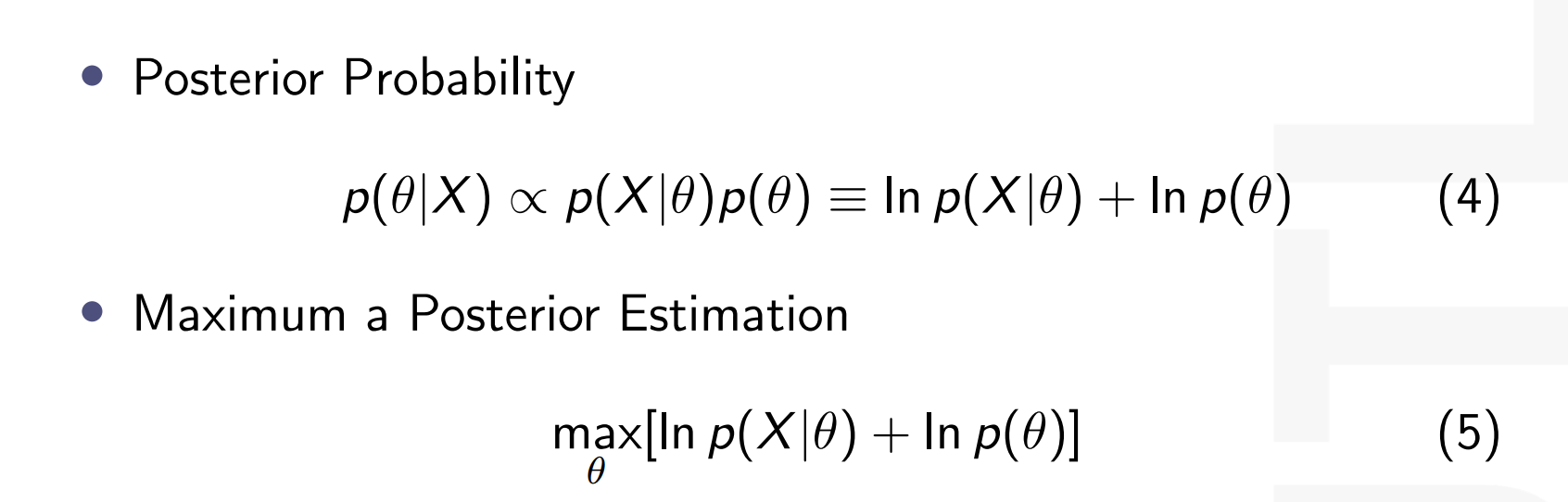
## Lecture 08(a)

**With data set and priors given, *pdf* of the data known, how to calculate the *pdf* ’s parameters based on MAP approach.**

MAP:Maximum a Posteriori Probability Estimation

Estimate:估计 parameter:参数

MAP就是MLE的基础上加上先验概率的对数求导

****

**differences and relations between MLE and MAP**

相同：

1.If p(θ) obeys the uniform distribution, both estimates yield identical results

2.Furthermore, for the case N → ∞, regardless of the values of the variances, the MAP estimates to the ML one. This is a more general result.

不同：

1. In ML estimation, we consider θ as an unknown parameter. In MAP estimation, we consider θ as a random variable.

2.MLE usually has better properties when the sample size is large, while MAP may be more sensitive to the selection of prior distributions, especially in the case of less data

3.the difference between the ML and MAP estimates lies in the involvement of p(θ) in the latter case.

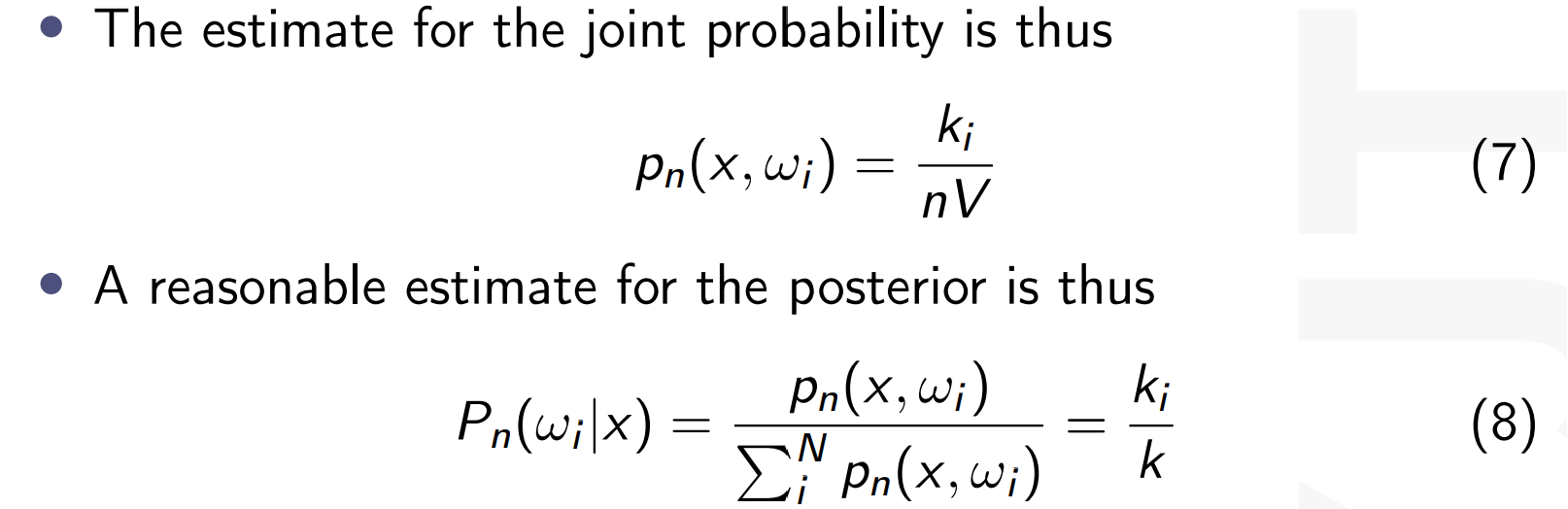
## Lecture 11

**k-NN Posterior Estimation**

非参数学习的分类算法，判别一个数据x为什么类别，就以x为中心，四周拉k个样本，以此判断新数据的类别。

k太小：受到个例影响Affected by individual cases

k太大：受到距离较远的特殊数据的影响Affected by special data that are far away

****

**以k个为一组，里面有ki的wi，种类wi的概率就是k个里面有ki的个数**

**Two parameters**

Number of Neighbors (k) and distance function

k:Number of Neighbors .

distance function:Euclidean or Manhattan

the value of k and the distance function

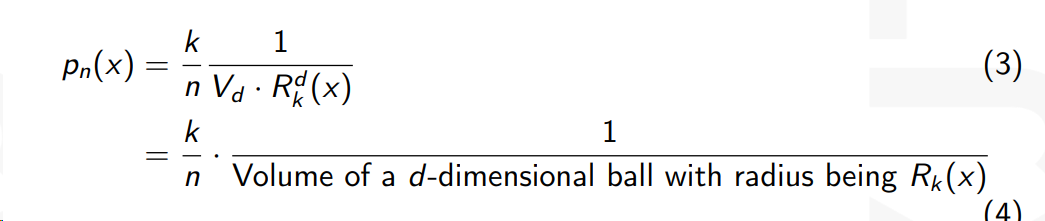
KNN优点：1.Extremely easy to chieve

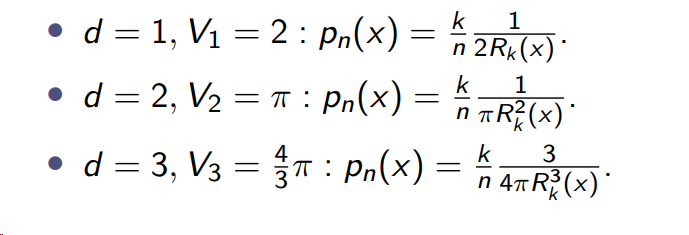
2.requires no training prior to making predictions. This makes KNN much faster

KNN缺点：1.KNN doesn’t work well with high dimensional data because it becomes difficult to calculate the distance.

2.KNN has high prediction cost for large datasets

3.KNN doesn’t work well with categorical features since it is difficult to find the distance.

**examples** 



## 其中k为选取的k，x为中心点，Rk（x）为样本点离x的距离

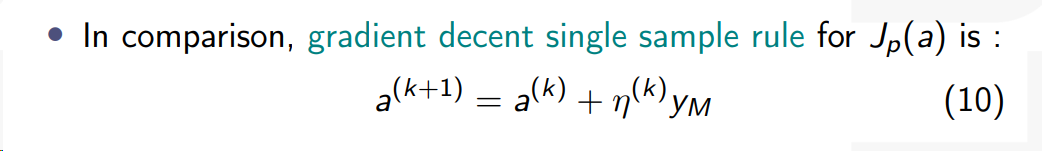
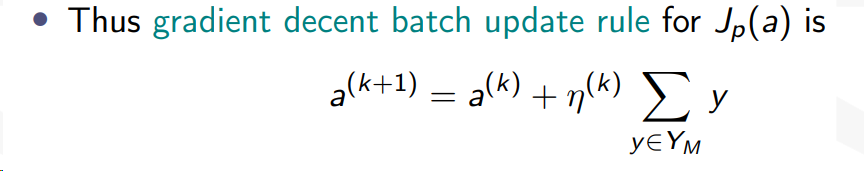
## Lecture 13

感知器：

输入一些特征，每个特征乘以相应的权重，最终决定相应的类别。

模型与真实数据之间存在一个损失函数，近似抛物线，最低点loss最小，模型越好。因此要不断梯度下降，直到最低点。

**Perceptron Example**



想象一个抛物线，η过大就导致下降过快，可能错过最低点

η过小就导致下降太慢，而且可能是极小值非最小值

数据加1，让数据升维度，从x-space到y-space

将数据归一化，即将不同类别的取负号放到同一类别

写出最终梯度更新公式

带入aty，若＞0，继续下一行，若小于0，则更新梯度

将最终aty带回g（y）并带回g（x）

## Lecture 14

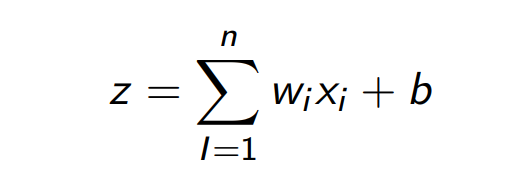
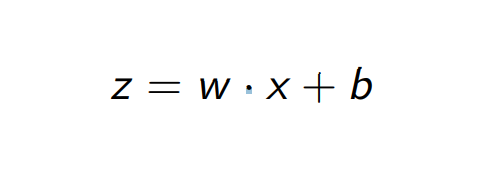
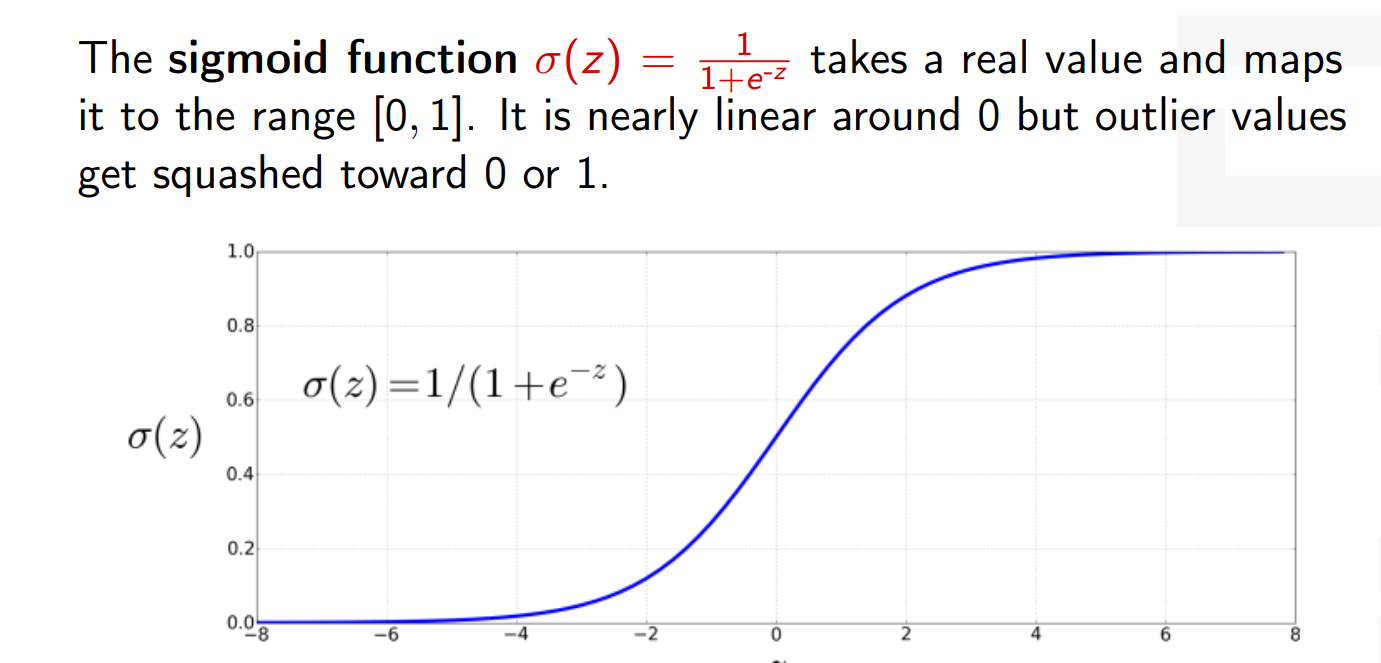
逻辑回归：将事情分为两类，或者判断多类中的一类

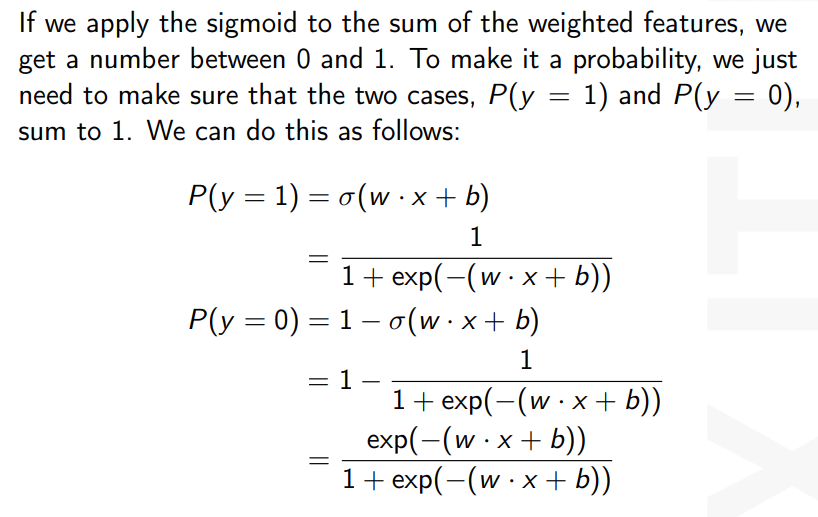
Linear 线性的

**sigmoid and softmax function (understand, not memorize)**

Sigmoid：减小极端影响，使分布集中

输出在【0，1】之间，代表概率

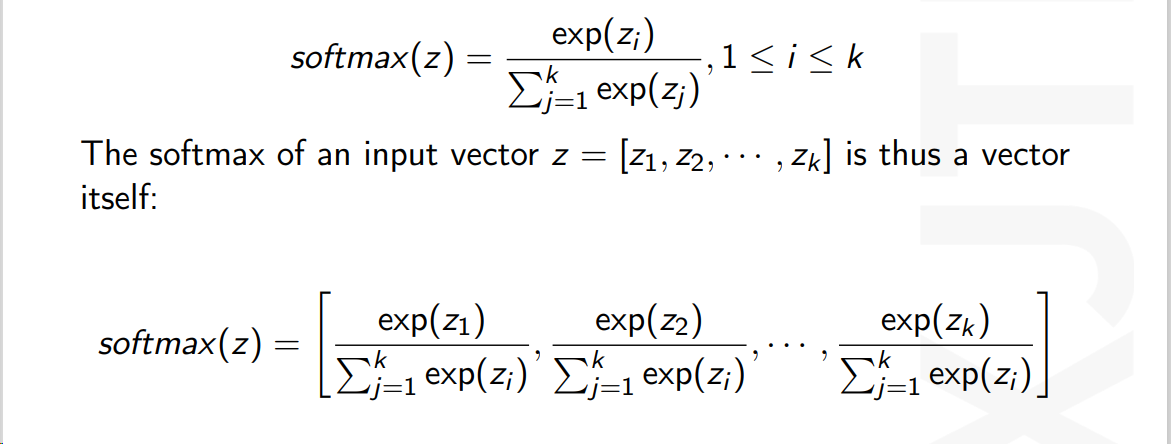
**** **** 

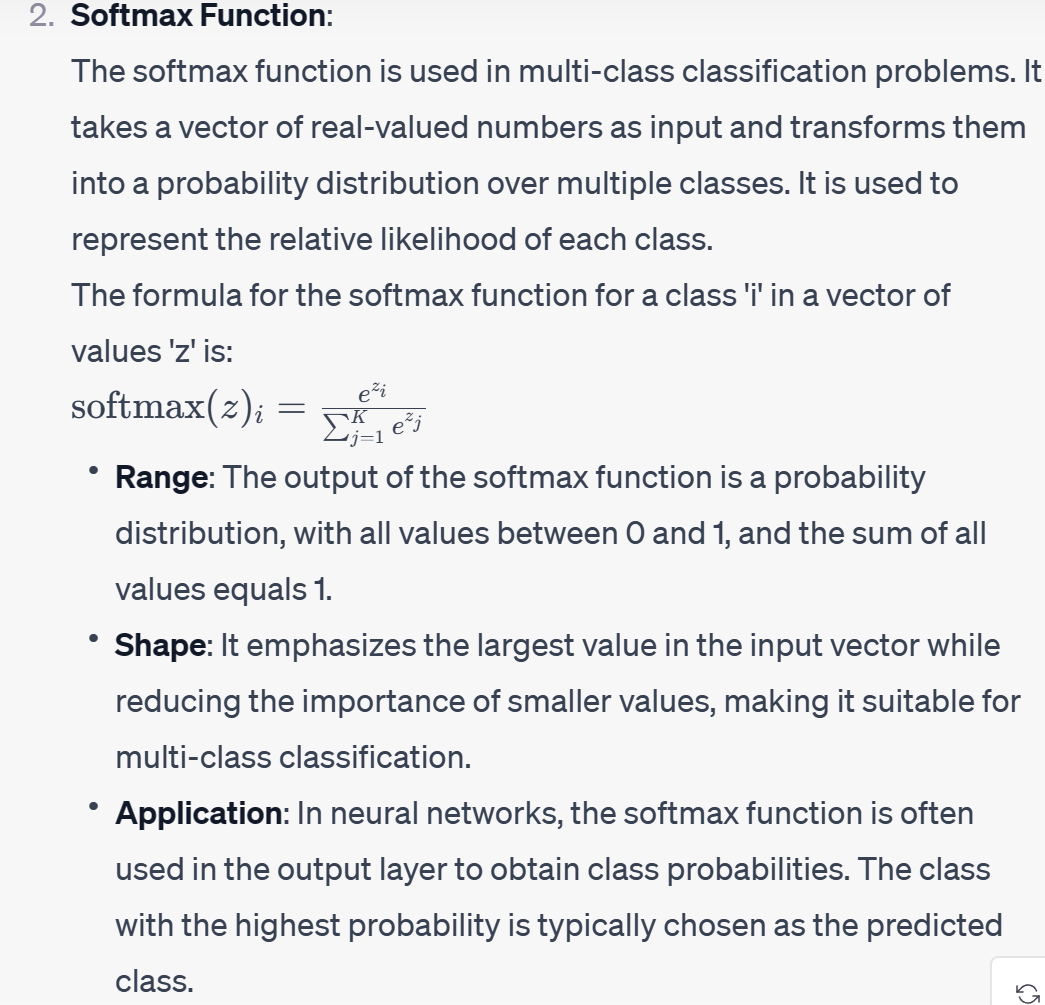
**** 

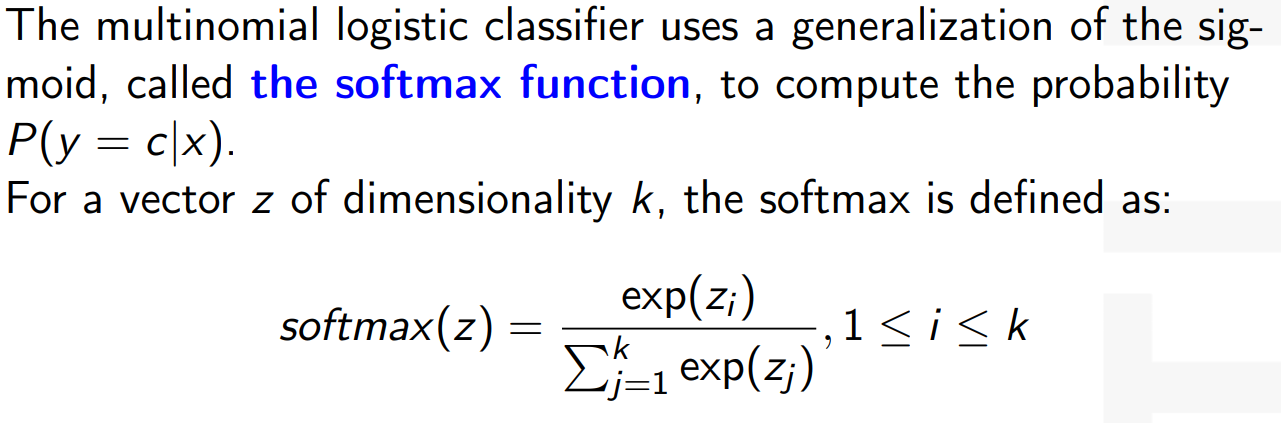
**Softmax：**

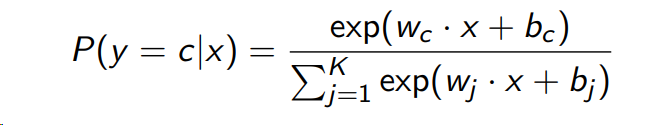
将多个类别相应的分数等比例压缩到合为1，表示为概率

在多分类问题中表示每个类别的相对可能性



****

****



## Lecture 15

**the idea behind SVM**

**1. 最大化间隔：**

**找到最优超平面：SVM 的目标是找到能够正确分类数据的超平面，同时确保不同类别之间的间隔（或距离）最大。**

**间隔：超平面与最近的数据点之间的距离称为间隔。最大化间隔有助于提高模型的泛化能力，减少过拟合的风险。**

**Maximizing Margin:**

**Optimal Hyperplane: The goal of SVM is to find a hyperplane that correctly classifies the data while ensuring that the margin, or distance, between different categories is maximized.**

**Margin: The distance between the hyperplane and the closest data point （the support data）is known as the margin. Maximizing the margin helps to enhance the model's generalization capabilities and reduce the risk of overfitting.**

**其距离为**

**2. 支持向量：**

**支持向量：距离超平面最近的数据点被称为支持向量。这些点对于定义超平面是关键的，因为它们决定了间隔的大小。**

**依赖于边界点：SVM 的决策仅依赖于支持向量，而非整个数据集。**

**Support Vectors:**

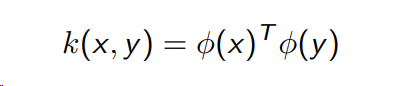
**Support Vectors: Data points that are closest to the hyperplane are referred to as support vectors. These points are crucial for defining the hyperplane as they determine the size of the margin.**

**Dependency on Boundary Points: The decision of SVM depends solely on the support vectors, not on the entire dataset**

**3. 核技巧：****当数据不是线性可分的时，SVM 使用核技巧将原始数据映射到更高维的空间（x变成**ϕ(x) **），但高维空间的内积不好计算，于是引入核函数k（x，z），从而隐式的在高维空间中计算内积**

**Handling Non-Linearly Separable Data:**

**Kernel Trick:**

**When the data is not linearly separable, SVM employs a kernel trick to map the original data to a higher-dimensional space (x becomes** **ϕ(x)). However, computing the inner product in this high-dimensional space can be computationally expensive. Therefore, a kernel function k(x,z) is introduced to facilitate the implicitly calculation of inner products in the high-dimensional space.**

**有效核指的是能满足**

**的式子。选择一个**

**ϕ(x)，使右边式子等于给出的左边式子**





Kernel Matrix is symmetric positive semidefinite = it is a valid Kernel Matrix

**4. 正则化参数C：**

**测量一定数量的错误分类，它控制最小化训练错误和控制模型复杂性之间的权衡。平衡错误样本数量与Margin宽度**

measuring some amount of misclassification,e it controls the trade-off between minimizing training errors and controlling model complexity.

**防止过拟合：通过引入正则化参数 C，SVM 控制模型复杂度，防止过拟合。**

**较小的 C 值会增加间隔的大小但容忍一些分类错误，而较大的 C 值则会减小间隔大小，追求在训练集上的完美分类。**

**Regularization:**

**Preventing Overfitting: By introducing the regularization parameter C, SVM controls the complexity of the model to prevent overfitting.**

**A smaller C value increases the margin size but tolerates some classification errors, while a larger C value decreases the margin size, striving for perfect classification on the training set.**

**concept of the kernel trick, use examples**

**how the regularization parameter *C* works in SVM**



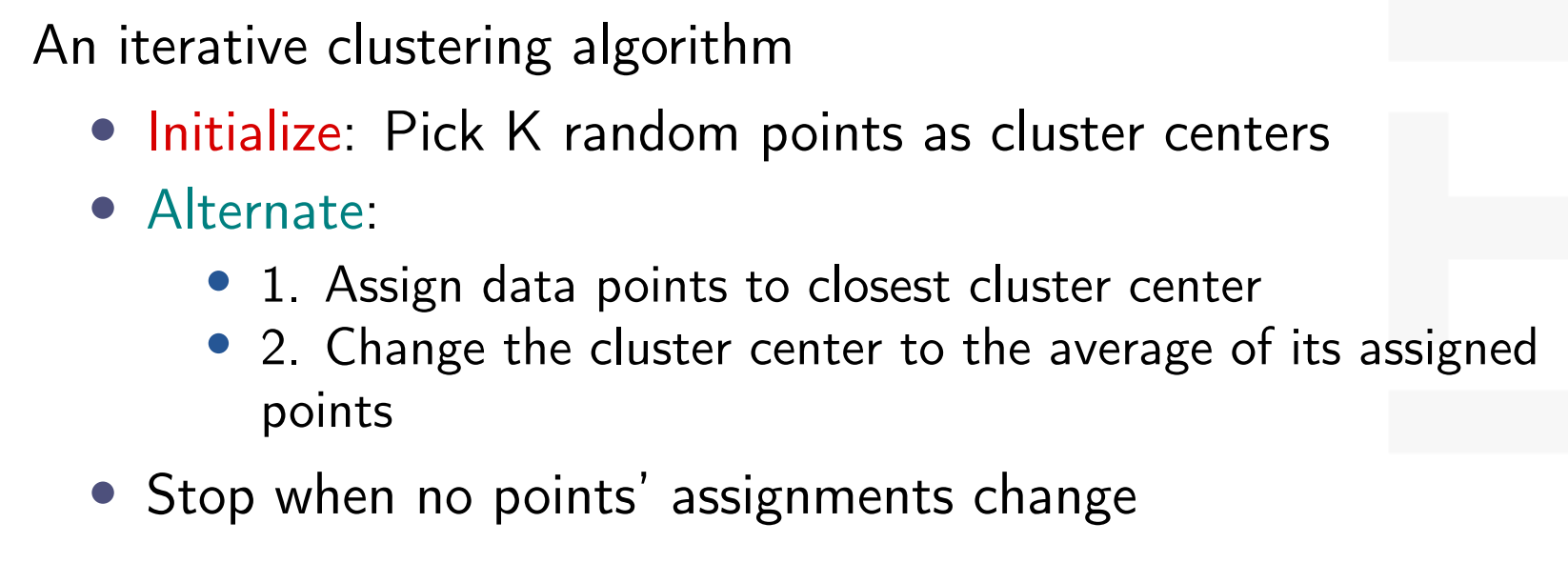
**SVM优点：**

**SVM缺点：**

**1、无法做线性不可分的数据**

## Lecture 16

* + **explain *K*-means**

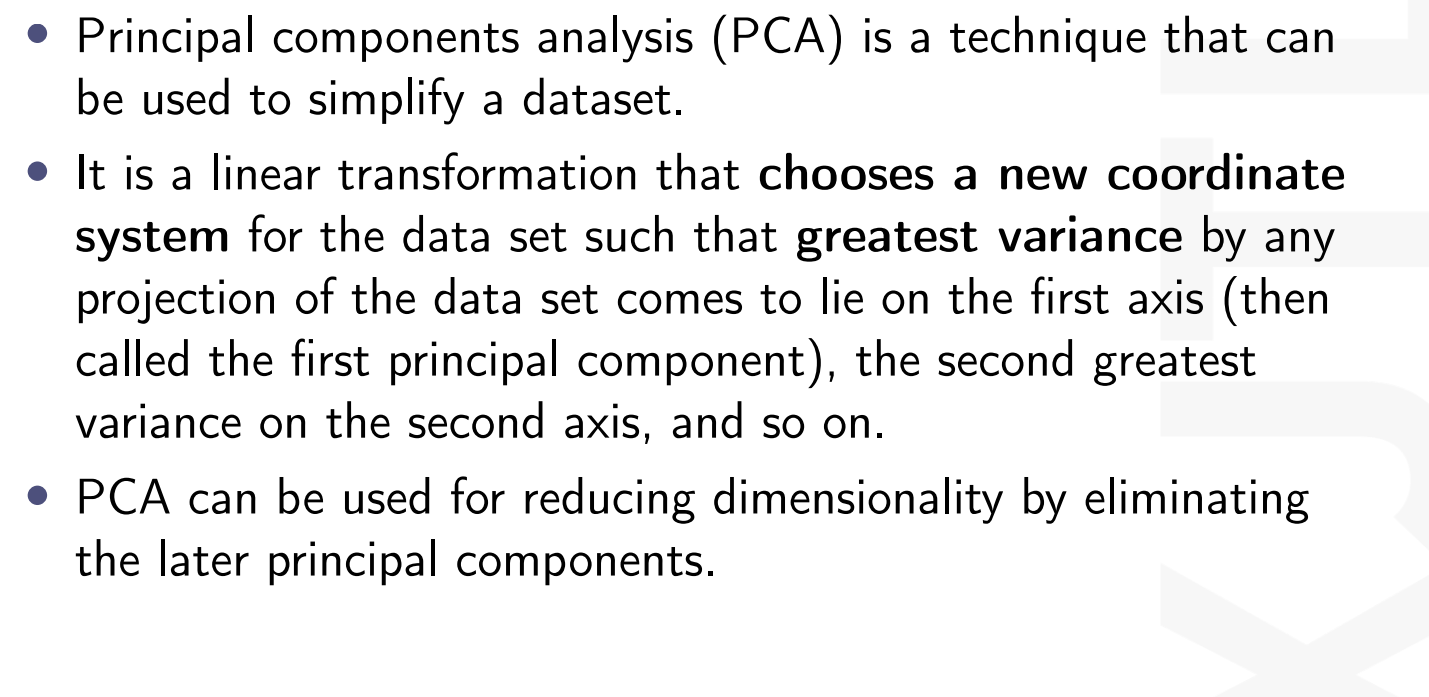
****

## **Lecture 17**

* + **explain the idea of PCA**

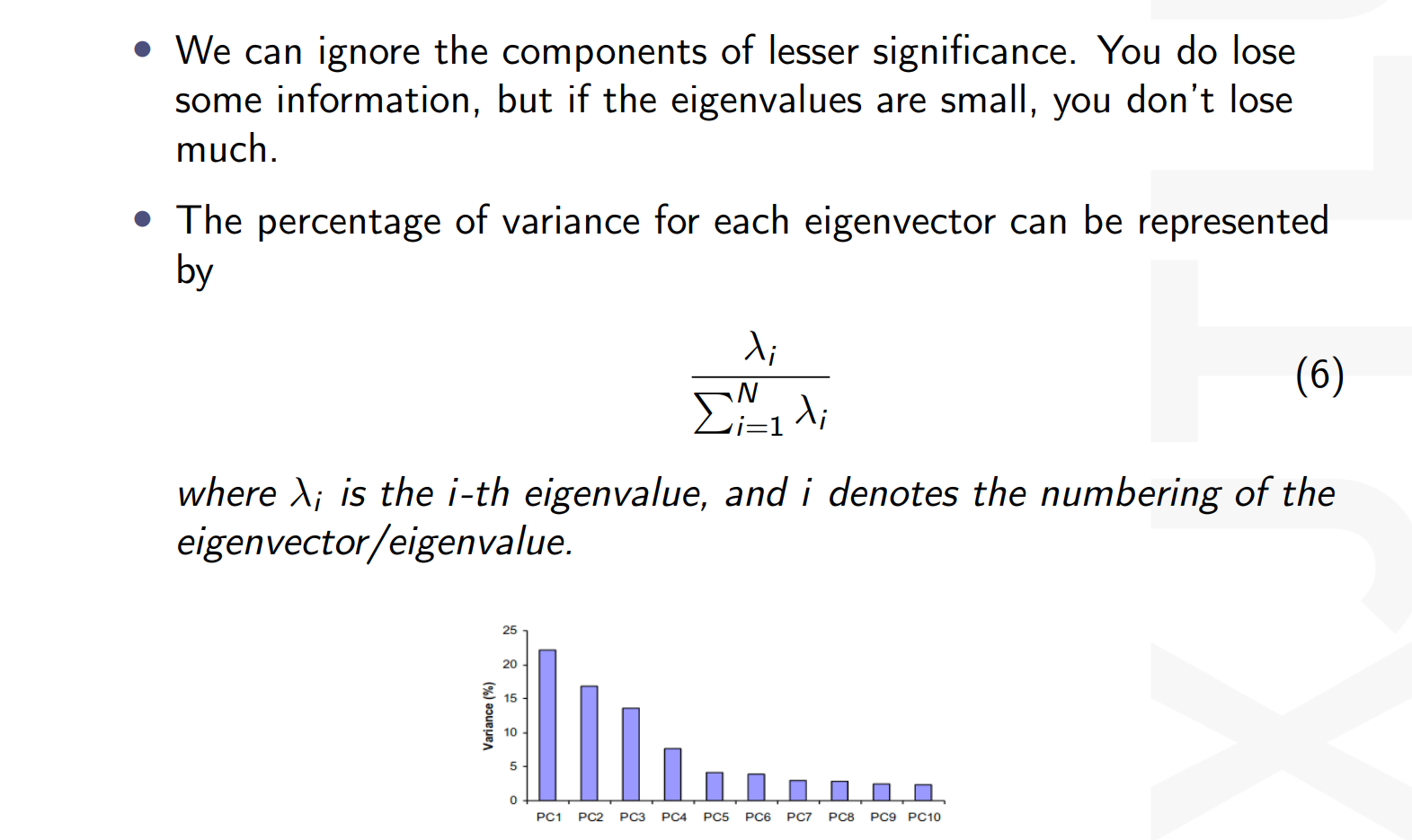
**Principal components analysis**

* + **PCA找坐标系，其中第一主成分是找一个坐标轴，后面同理**

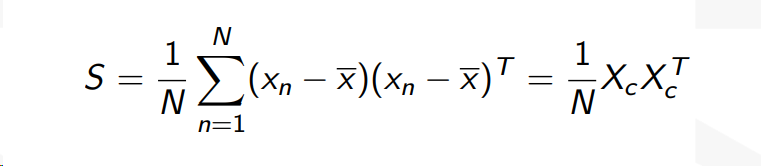
****

Our goal is to project the data onto a space having dimensionality M< D while maximizing the variance of the projected data.

It is a linear transformation that chooses a new coordinate system for the data set such that greatest variance by any projection of the data set comes to lie on the first axis (then called the first principal component), the second greatest variance on the second axis, and so on. • PCA can be used for reducing dimensionality by eliminating the later principal components.

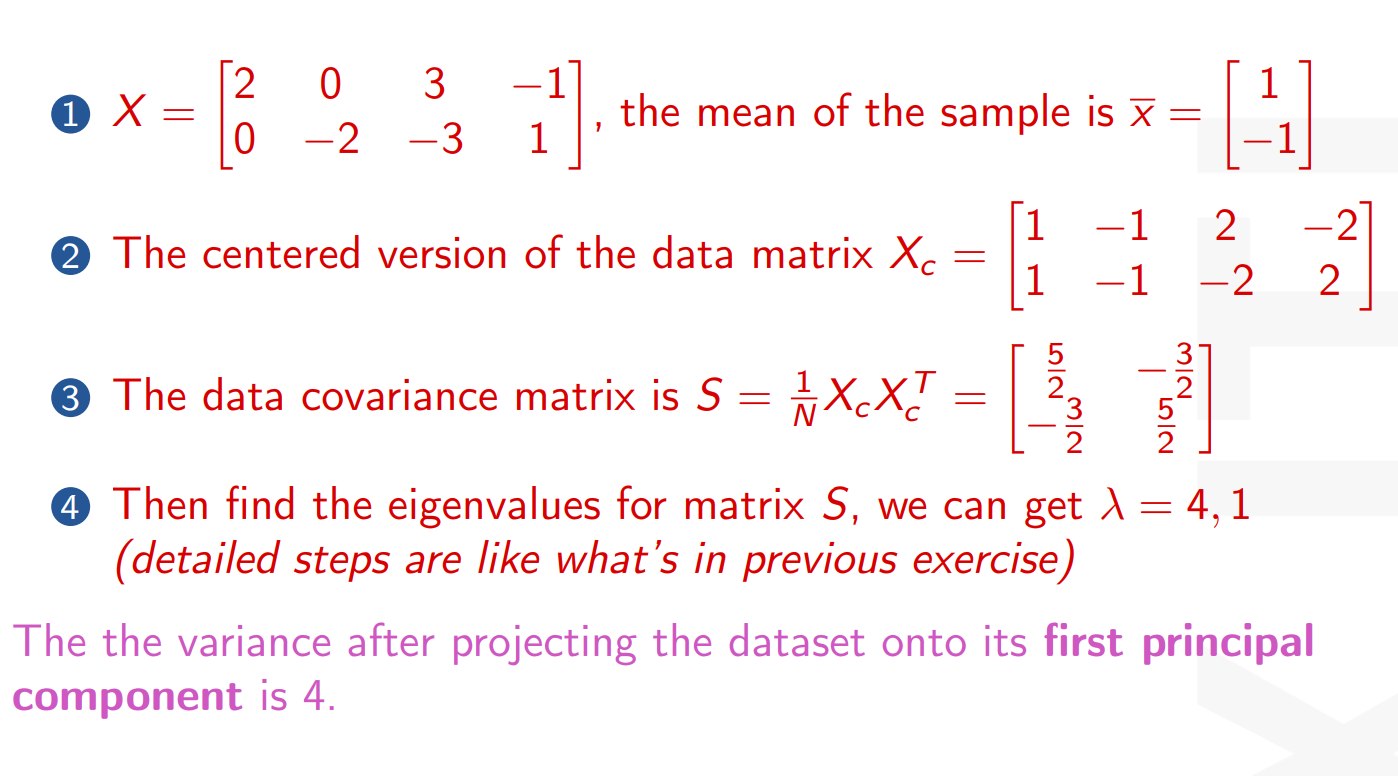
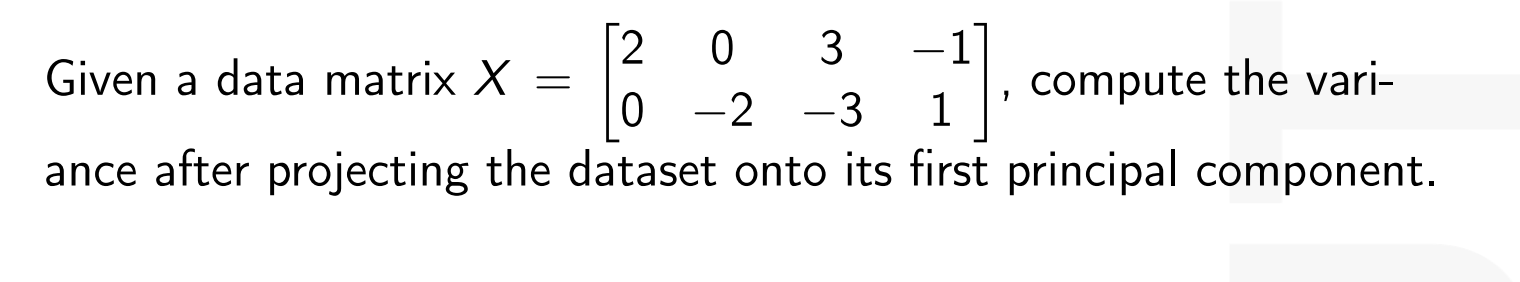
* + **role of eigenvalues and eigenmatrix**
  + Data dimensionality reduction: In data analysis, eigenvalues and eigenvectors are used in principal component analysis (PCA) to reduce the dimensionality of the data by finding the direction of the greatest change in the data (eigenvectors), and the corresponding eigenvalues represent the degree of change in each direction.
  + **how many PCs should be retained**
  + ****

**Exercises**



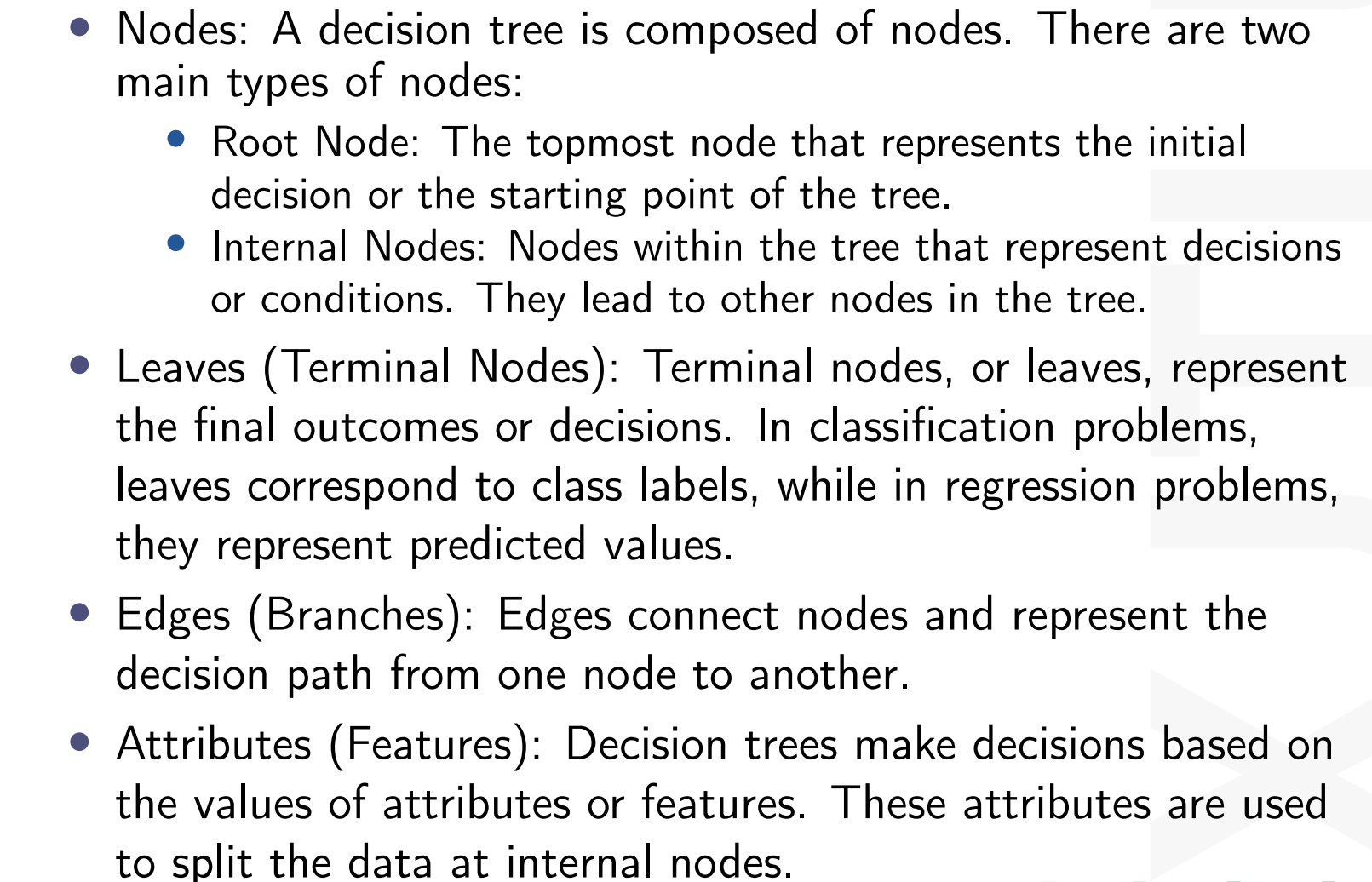
**求PCA投影后的主成分步骤：**

1. **算出每行的均值**
2. **写出Xc**
3. **写出S（data covariance matrix）**
4. **算出eigenvalue and eigenmatrix**

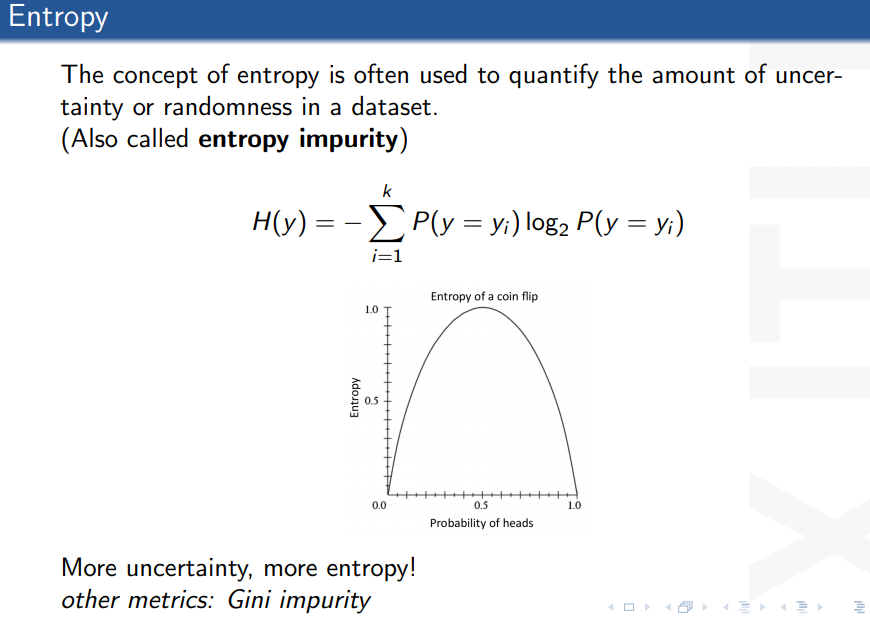
****

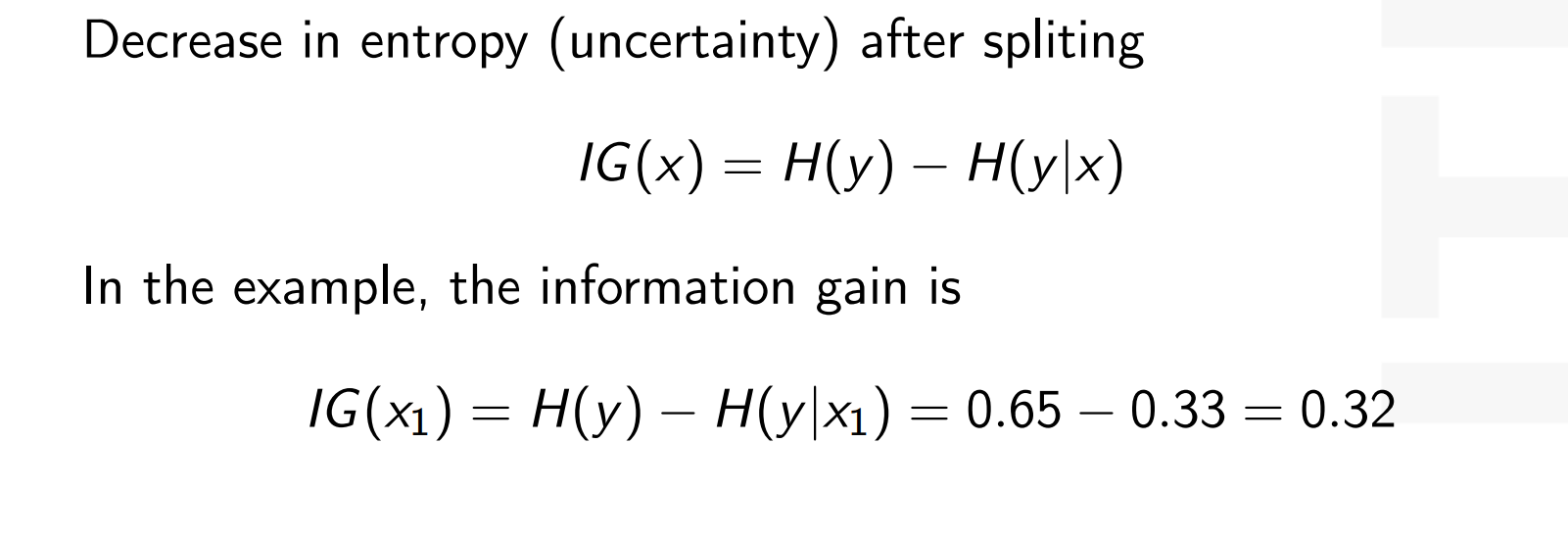
## Lecture 19

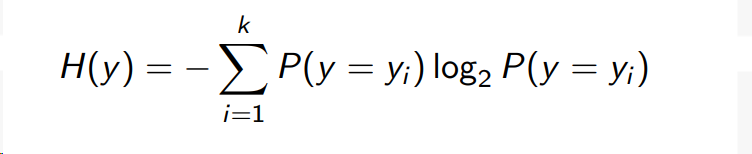
* + **Decision Trees construction**

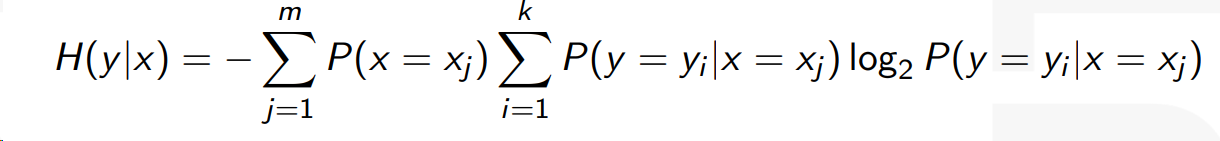
****

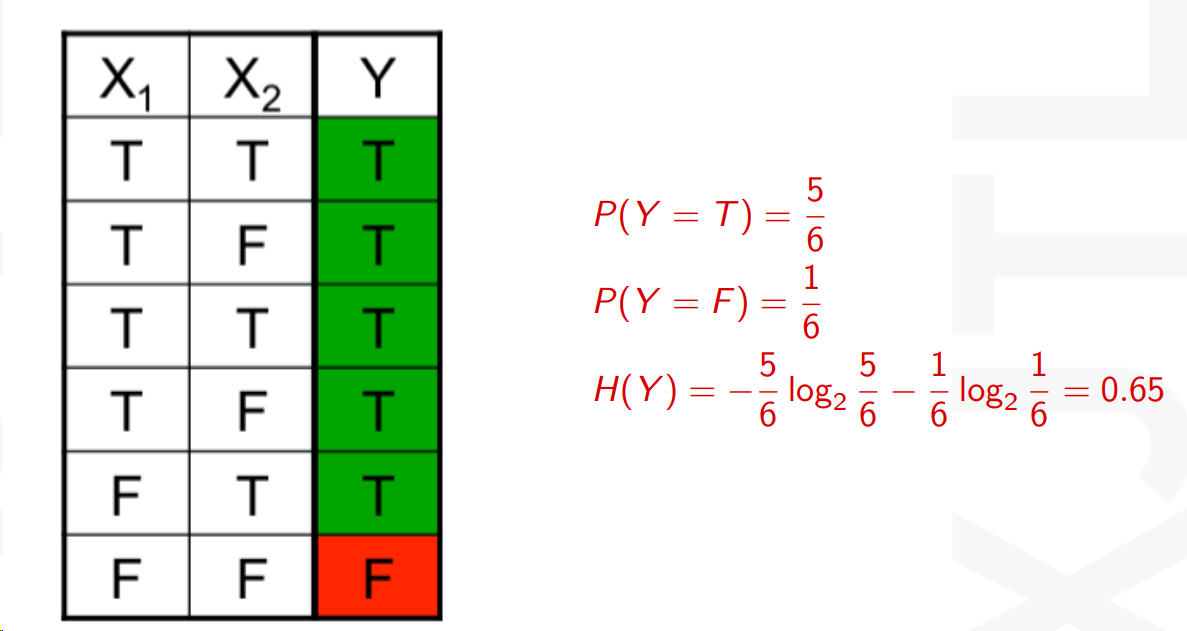
* + **Attributes指的是branches当中的特征**
  + **Information gain**

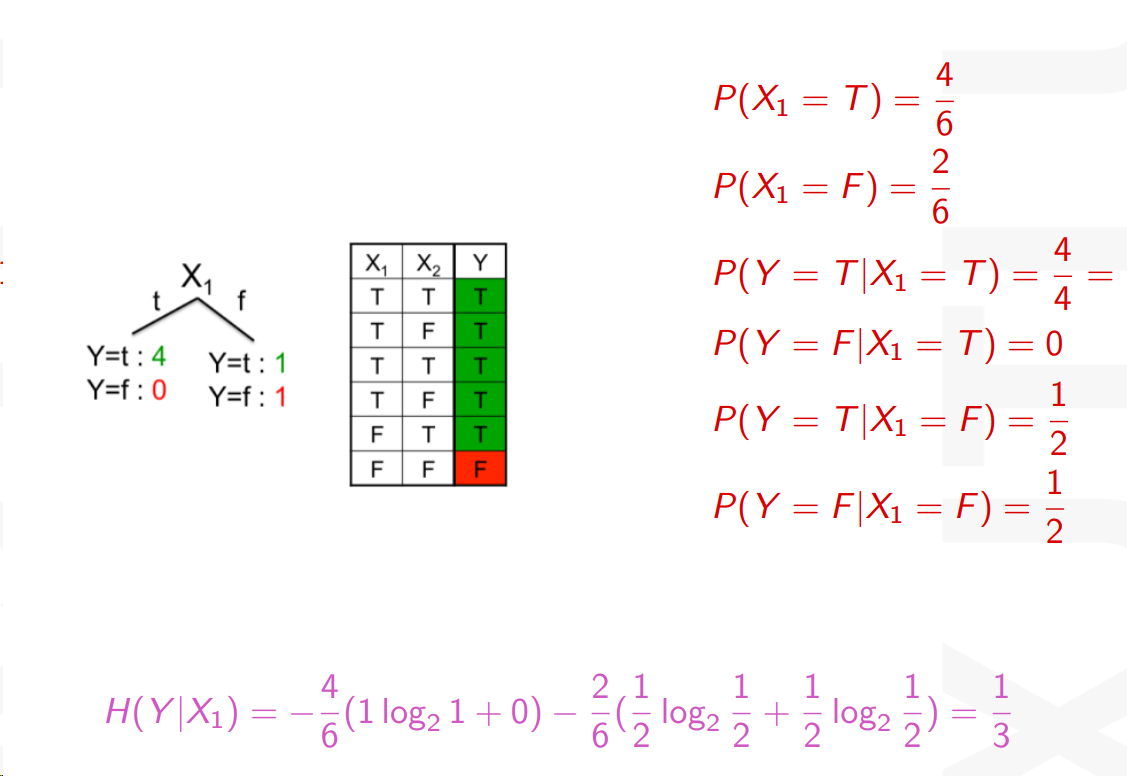
****

****



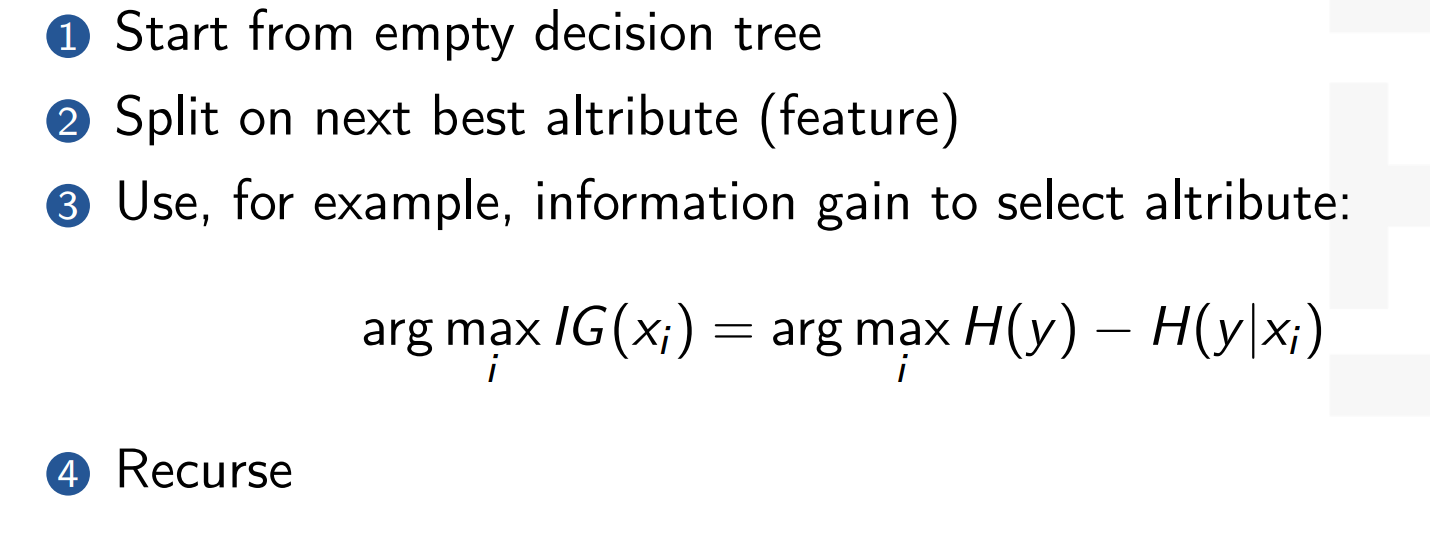






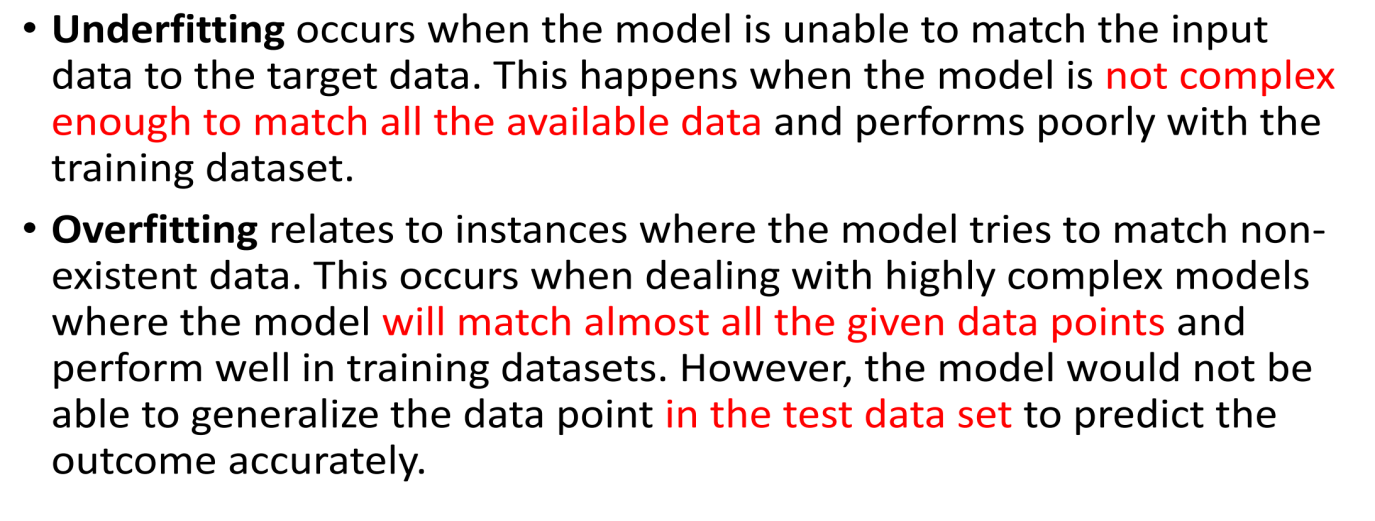
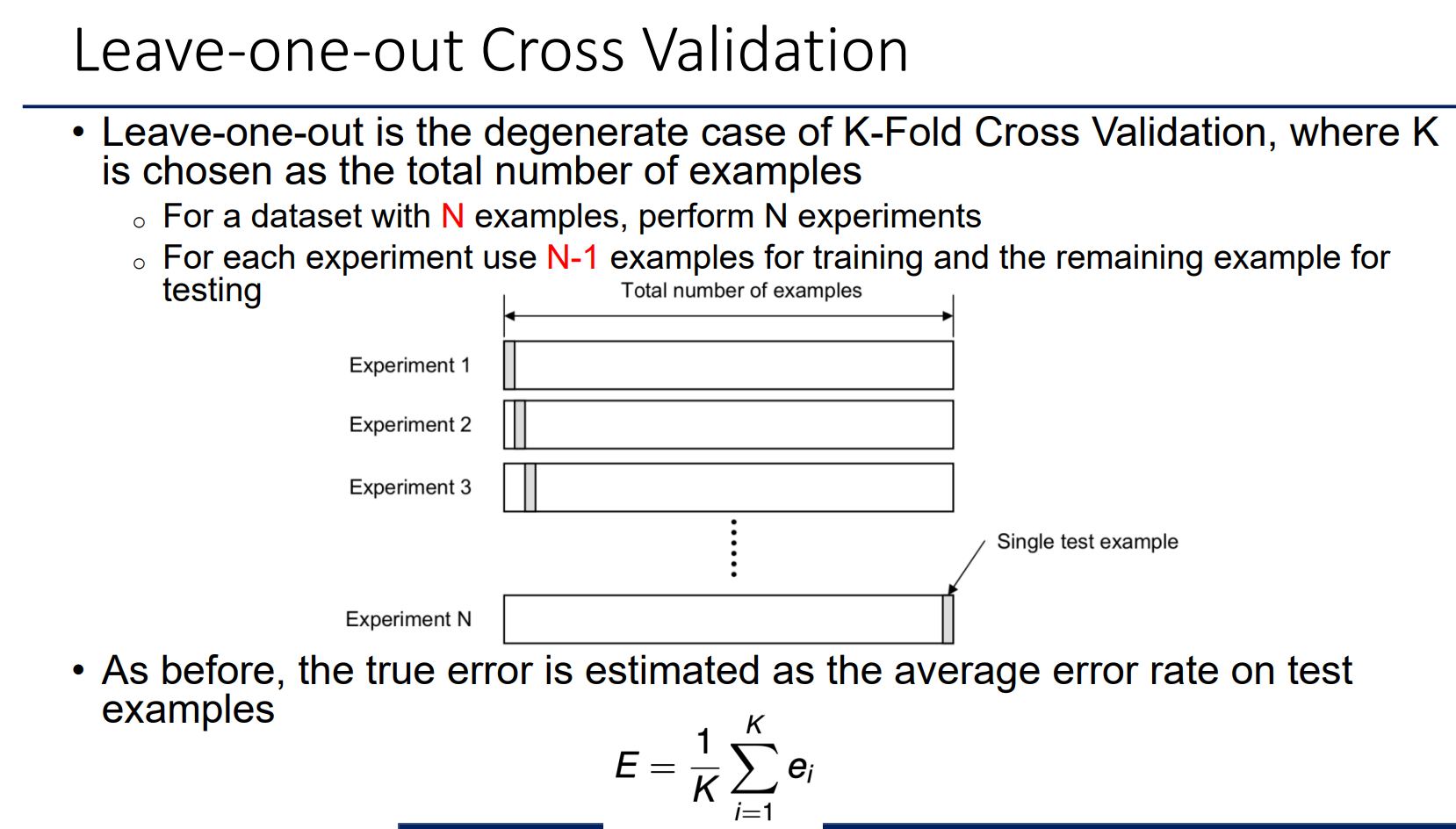
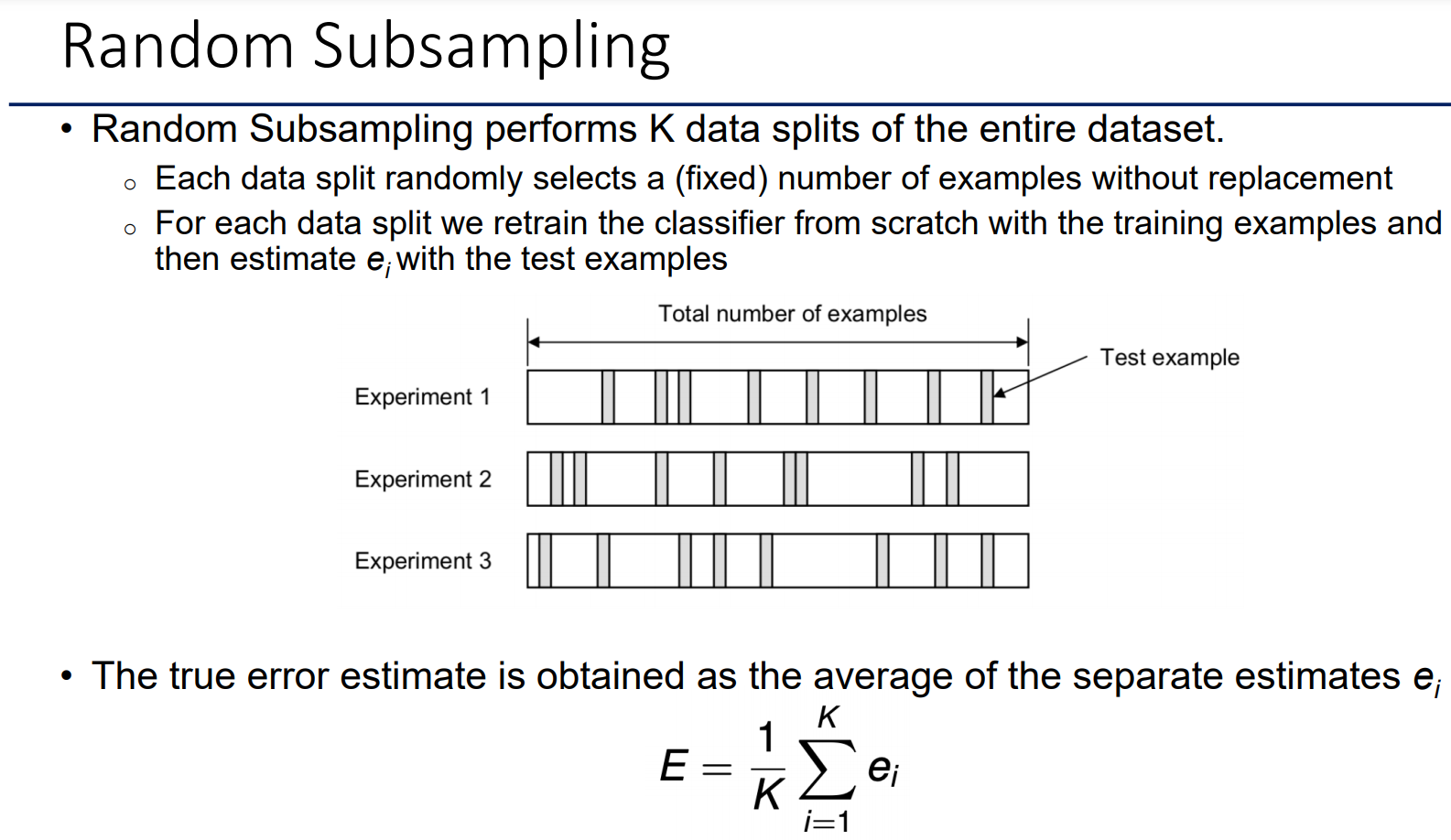
信息增益 = 熵（原先不确定性）- 条件熵（给定一个信息后的不确定性）

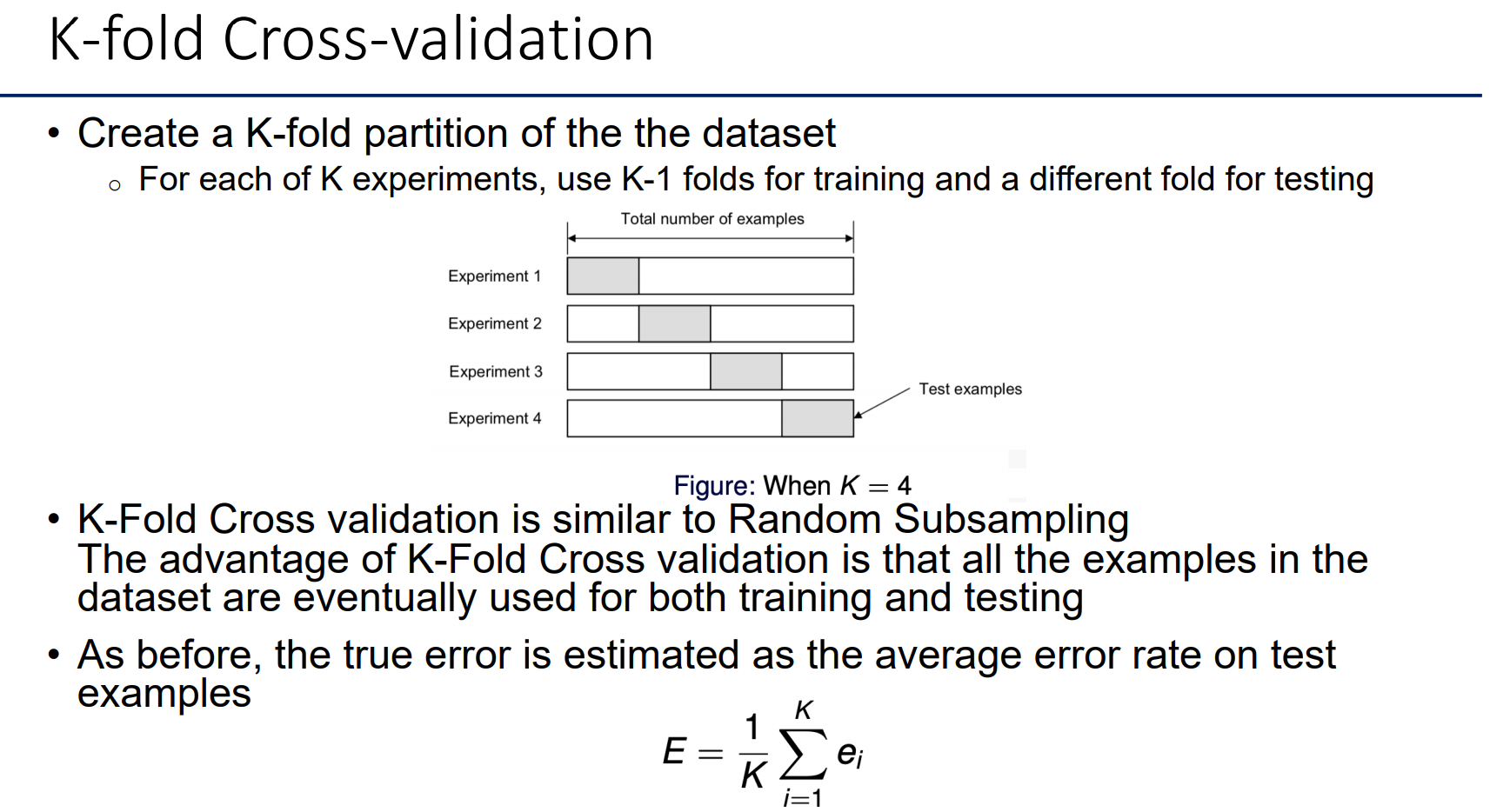
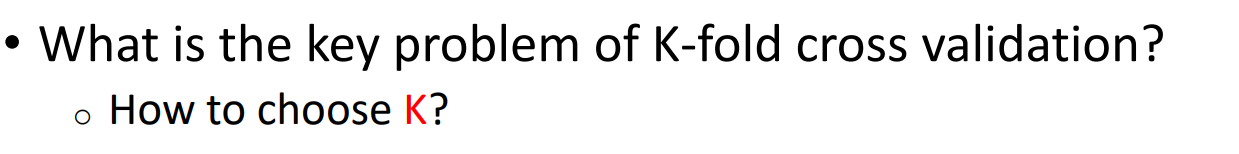
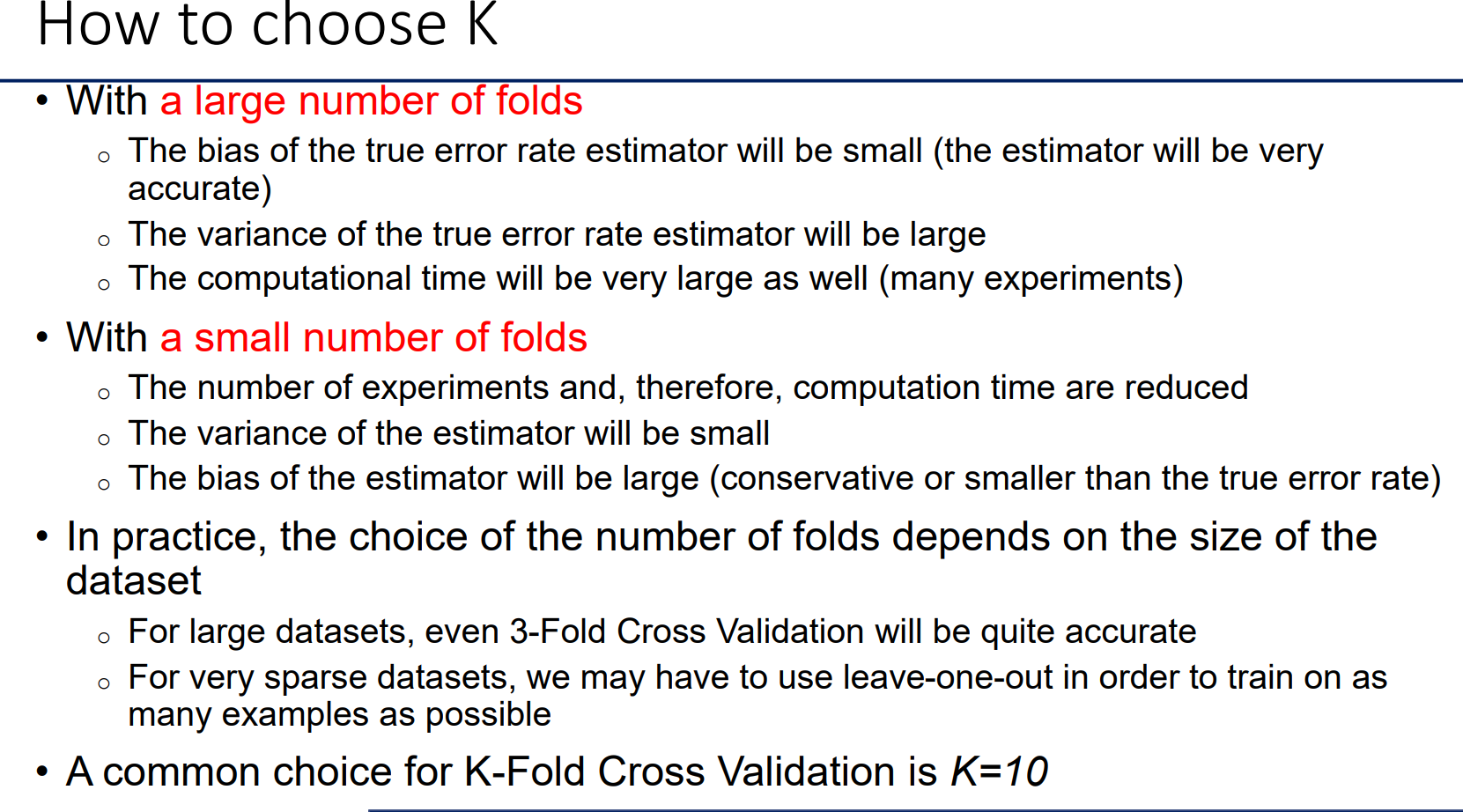
Imformation = Original uncertainty - Uncertainty given a piece of information

****

## Lecture 20

* + **Underfitting and Overfitting**

**cross validation**

****  

# Distribution

* + **The exam paper consists only of short-answer questions; there are no multiple-choice questions or no fill-in-the blank questions.**
  + **There are 5 primary questions in total, each with 2-4 sub-questions.**
  + **The 1st question is calculation-based.**
  + **The 2nd question is both calculation-based and descriptive/explanatory.**
  + **Question 3,4 and 5 are only descriptive/explanatory.**

# How to use mock exam paper

**The mock exam paper does reflect**

* + **the number of questions**
  + **types of questions**
  + **level of difficulty**

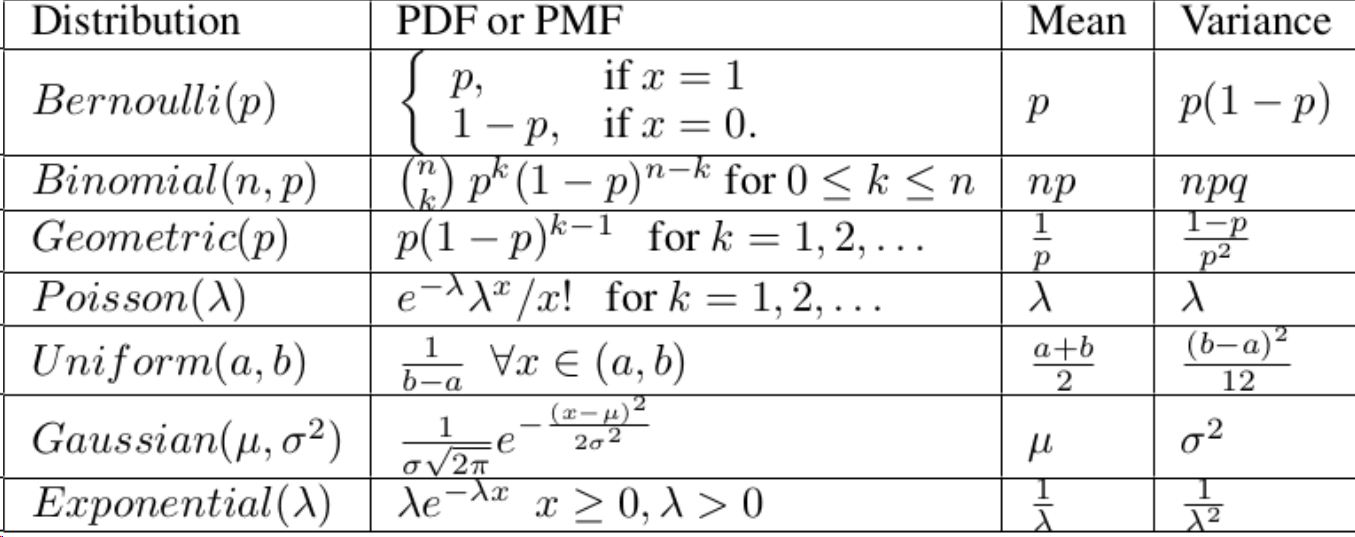
**The mock exam paper does NOT reflect**

* + **content of the final exam**

**mock exam paper release date: Oct.27th 23:00**

**课外**

**对角协方差矩阵 diagonal covariance matrix**



**Label指的就是种类**

**特征指的就是数据的特性（大小、颜色）**