

# 第一章 绪论

## 统计学习

总目标：考虑学习什么样的模型和如何学习模型，使得模型能对数据进行准确的预测与分析，同时尽可能提高学习效率。

统计学习的步骤
1. 得到一个有限的训练数据集合
2. 确定包含所有可能的模型的假设空间，即学习模型的集合
3. 确定模型选择的准则，即学习的策略
4. 确定模型选择的准则，即学习的策略
5. 通过学习方法选择最优的模型
6. 利用学习的最优模型对新数据进行预测或分析

## 分类

### 基本分类

监督学习，样本表示 $(x, y)$ 对，在无监督学习里面，样本就是 $x$ 。

- 监督学习
- 无监督学习
- 强化学习

### 按模型分类

- 概率模型
- 非概率模型

在监督学习中，概率模型是生成模型，非概率模型是判别模型。

### 按算法分类

- 在线学习
- 批量学习

在线学习通常比批量学习更难。

### 按技巧分类

- 贝叶斯学习
- 核方法

# 基本概念

- 1. 输入空间、特征空间、输出空间
- 2. 联合概率分布： 监督学习假设输入输出X和Y遵循联合概率分布P(X,Y)， 这是基本假设。
- 3. 假设空间

## 统计学习方法三要素

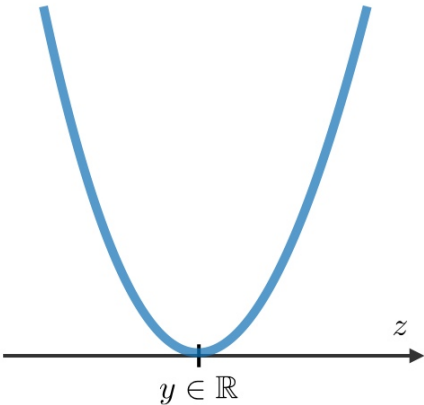
### 1. 模型

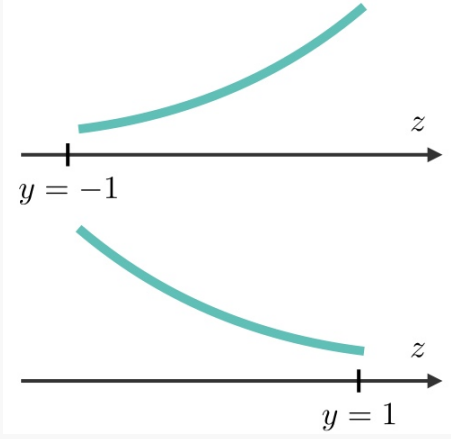
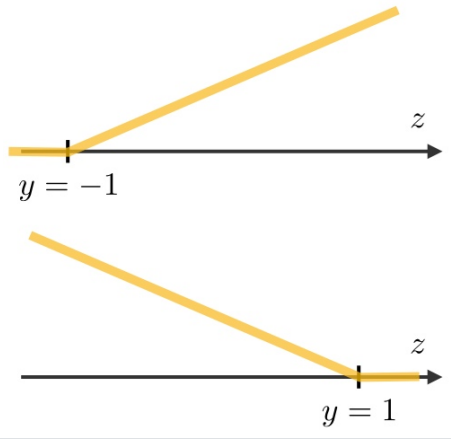
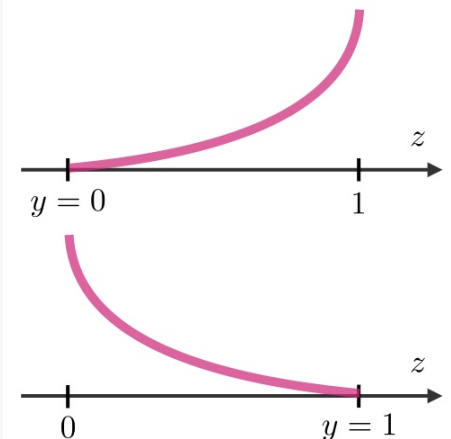
	假设空间 $\mathcal{F}$	输入空间 $\mathcal{X}$	输出空间 $\mathcal{Y}$	参数空间
决策函数	$\mathcal{F} = \{f_{\theta}   Y = f_{\theta}(x), \theta \in \mathbf{R}^n\}$	变量	变量	$\mathbf{R}^n$
条件概率分布	$\mathcal{F} = \{P   P_{\theta}(Y X), \theta \in \mathbf{R}^n\}$	随机变量	随机变量	$\mathbf{R}^n$

### 2. 策略（Loss）

#### 损失函数（loss function）

Loss function – A loss function is a function  $L : (z,y) \in \mathbf{R} \times Y \rightarrow L(z,y) \in \mathbf{R}$  that takes as inputs the **predicted value z** corresponding to the **real data value y** and outputs how different they are. The common loss functions are summed up in the table below:

Loss Function	公式	图例	应用
Least squared	$L(Y, f(X)) = \frac{1}{2}(Y - f(X))^2$		Linear regression

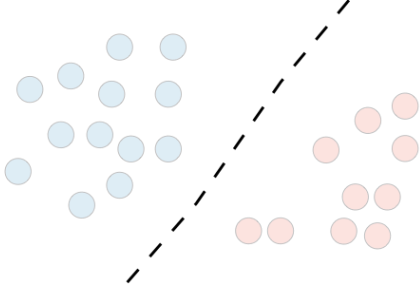
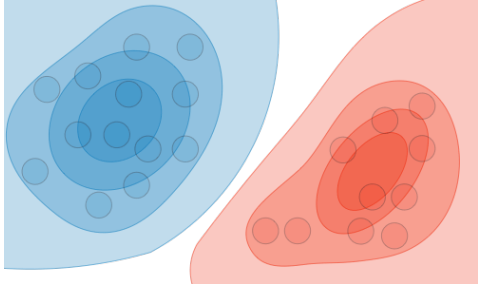
Logistic	$\log(1 + \exp(-Y * f(X)))$		Logistic regression
Hinge	$\max(0, 1 - Y * f(X))$		SVM
Cross-entropy	$-Y * \log(f(X)) + (1 - Y) * \log(1 - f(X))$		Neural Network

代价函数 (cost function)

### 3. 算法 (优化算法)

4.

### 生成模型与判别模型

	<b>Discriminative model</b>	<b>Generative model</b>
Goal	Directly estimate $P(y x)$	Estimate $P(x y)$ to deduce $P(y x)$
What's learned	Decision boundary	Probability distributions of the data
Illustration		
Examples	Regressions, SVMs	GDA, Naive Bayes