贝叶斯理论导读

SYSU_CIS@lindayong

有监督学习

N个样本点构成的数据集，其中第 i 个样本表示为 $(\boldsymbol{x}_i, y_i)$。

$\boldsymbol{x}_i$：第 i 个样本的特征值　　y_i：第 i 个样本的标签

Living area (feet^2)	#bedrooms	Price (1000$s)
2104	3	400
1600	3	330
2400	3	369
1416	2	232
3000	4	540
⋮	⋮	⋮

缺一张表示数据集的图片

传统模型

1. 构建模型
   - $\hat{y}_i = f(\boldsymbol{x}_i; \boldsymbol{\theta})$

给定模型的参数 $\boldsymbol{\theta}$，对于某个特征值 $\boldsymbol{x}_i$，模型的预测结果为 y_i 。

Living area (feet^2)	#bedrooms	Price (1000$s)
2104	3	400
1600	3	330
2400	3	369
1416	2	232
3000	4	540
⋮	⋮	⋮

$$\hat{y}_i = \theta_1 x_{i1} + \theta_2 x_{i2}$$

传统模型

1. 构建模型

$$\hat{y}_i = f(\boldsymbol{x}_i; \boldsymbol{\theta})$$

给定模型的参数 $\boldsymbol{\theta}$，对于某个特征值 $\boldsymbol{x}_i$，模型的预测结果为 y_i 。

2. 训练模型

- 2.1 样本的损失函数 $\mathrm{loss}(\mathrm{y_i}, \hat{y}_i; \boldsymbol{\theta})$
- 2.2 数据集的损失函数

$$\mathrm{loss}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{loss}(y_i, \hat{y}_i; \boldsymbol{\theta})$$

- 2.3 训练目标 $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \ \mathrm{loss}(\boldsymbol{\theta})$

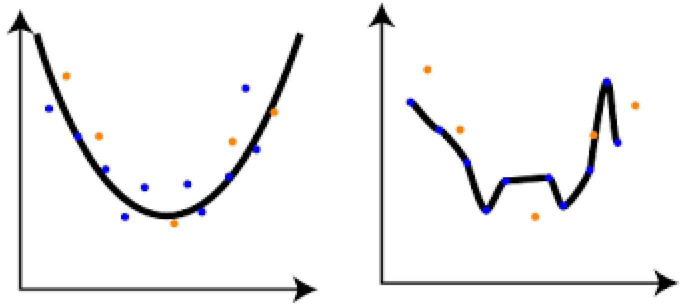
传统模型

1. 构建模型

$$\hat{y}_i = f(\boldsymbol{x}_i;\boldsymbol{\theta})$$

给定模型的参数 $\boldsymbol{\theta}$，对于某个特征值 $\boldsymbol{x}_i$，模型的预测结果为 y_i 。

2. 训练模型



- 2.1 样本的损失函数 $\mathrm{loss}(\mathrm{y_i}, \hat{y}_i; \boldsymbol{\theta})$
- 2.2 数据集的损失函数

$$\mathrm{loss}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{loss}(y_i, \hat{y}_i; \boldsymbol{\theta}) + \mathrm{reg}(\boldsymbol{\theta})$$

- 2.3 训练目标 $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathrm{loss}(\boldsymbol{\theta})$

3. 模型预测

$$\hat{y}_i = f(\boldsymbol{x}_i;\hat{\boldsymbol{\theta}})$$

传统模型

1. 构建模型

$$\hat{y}_i = f(\boldsymbol{x}_i;\boldsymbol{\theta})$$

2. 训练模型

- 2.1 样本的损失函数 $\mathrm{loss}(\mathrm{y_i}, \hat{y}_i; \boldsymbol{\theta})$
- 2.2 数据集的损失函数

$$\mathrm{loss}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{loss}(y_i, \hat{y}_i; \boldsymbol{\theta}) + \mathrm{reg}(\boldsymbol{\theta})$$

- 2.3 训练目标 $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \ \mathrm{loss}(\boldsymbol{\theta})$

3. 模型预测

$$\hat{y}_i = f(\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}})$$

对于其他预测结果 y_i^* 靠谱程度的衡量：

如果 y_i 存在，$\mathrm{loss}(y_i, y_i^*; \boldsymbol{\theta})$

传统模型

1. 构建模型

$$\hat{y}_i = f(\boldsymbol{x}_i; \boldsymbol{\theta})$$

2. 训练模型

- 2.1 样本的损失函数 $\mathrm{loss}(\mathrm{y_i}, \hat{y}_i; \boldsymbol{\theta})$
- 2.2 数据集的损失函数

$$\mathrm{loss}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{loss}(y_i, \hat{y}_i; \boldsymbol{\theta}) + \mathrm{reg}(\boldsymbol{\theta})$$

- 2.3 训练目标 $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathrm{loss}(\boldsymbol{\theta})$

3. 模型预测

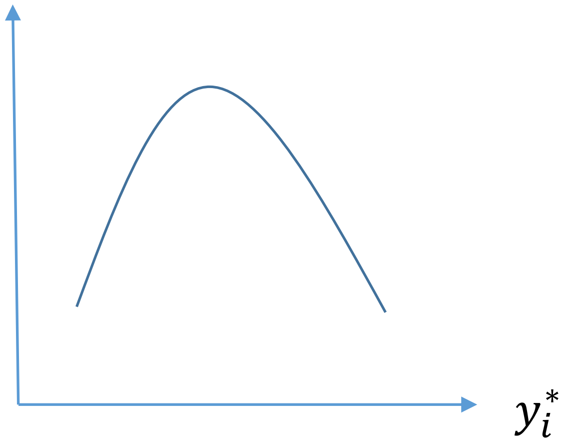
$$\hat{y}_i = f(\boldsymbol{x}_i; \widehat{\boldsymbol{\theta}})$$

概率

贝叶斯模型

1. 构建模型

$$p(y_i^* | \boldsymbol{x}_i; \boldsymbol{\theta})$$



传统模型

1. 构建模型

$$\hat{y}_i = f(\boldsymbol{x}_i; \boldsymbol{\theta})$$

2. 训练模型

- 2.1 样本的损失函数 $\mathrm{loss}(\mathrm{y_i}, \hat{y}_i; \boldsymbol{\theta})$
- 2.2 数据集的损失函数

$$\mathrm{loss}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{loss}(y_i, \hat{y}_i; \boldsymbol{\theta}) + \mathrm{reg}(\boldsymbol{\theta})$$

- 2.3 训练目标 $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathrm{loss}(\boldsymbol{\theta})$

3. 模型预测

$$\hat{y}_i = f(\boldsymbol{x}_i; \hat{\boldsymbol{\theta}})$$

概率

贝叶斯模型

1. 构建模型

$$p(y_i^* | \boldsymbol{x}_i; \boldsymbol{\theta})$$

2. 训练模型

- 2.2 数据集的似然函数

$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N} p(y_i|\boldsymbol{x}_i; \boldsymbol{\theta})$$

- 2.3 训练目标 $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} -p(\mathcal{D}|\boldsymbol{\theta})$

传统模型

1. 构建模型

$$\hat{y}_i = f(\boldsymbol{x}_i; \boldsymbol{\theta})$$

2. 训练模型

- 2.1 样本的损失函数 $\mathrm{loss}(\mathrm{y_i}, \hat{y}_i; \boldsymbol{\theta})$
- 2.2 数据集的损失函数

$$\mathrm{loss}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{loss}(y_i, \hat{y}_i; \boldsymbol{\theta}) + \mathrm{reg}(\boldsymbol{\theta})$$

- 2.3 训练目标 $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathrm{loss}(\boldsymbol{\theta})$

3. 模型预测

$$\hat{y}_i = f(\boldsymbol{x}_i; \hat{\boldsymbol{\theta}})$$

概率

贝叶斯模型

1. 构建模型

$$p(y_i^* | \boldsymbol{x}_i; \boldsymbol{\theta})$$

2. 训练模型

- 2.2 数据集的似然函数

$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N} p(y_i | \boldsymbol{x}_i; \boldsymbol{\theta})$$

贝叶斯公式

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta) * p(\theta)$$

- 2.3 训练目标 $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} -p(\mathcal{D}|\boldsymbol{\theta})$

传统模型

1. 构建模型

$$\hat{y}_i = f(\boldsymbol{x}_i;\boldsymbol{\theta})$$

2. 训练模型

- 2.1 样本的损失函数 $\mathrm{loss}(\mathrm{y_i}, \hat{y}_i;\boldsymbol{\theta})$
- 2.2 数据集的损失函数

$$\mathrm{loss}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{loss}(y_i, \hat{y}_i;\boldsymbol{\theta}) + \mathrm{reg}(\boldsymbol{\theta})$$

- 2.3 训练目标 $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathrm{loss}(\boldsymbol{\theta})$

3. 模型预测

$$\hat{y}_i = f(\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}})$$

概率

点估计

贝叶斯估计

贝叶斯模型

1. 构建模型

$$p(y_i^*|\boldsymbol{x}_i;\boldsymbol{\theta})$$

2. 训练模型

- 2.2 参数的后验分布

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta) * p(\theta)$$

- 2.3 训练目标 $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} -p(\theta|\mathcal{D})$

3. 模型预测

$$\hat{y}_i = \arg\max_{y_i^*} p(y_i^*|\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}})$$

$$p(y_i^*|\boldsymbol{x}_i;\mathcal{D}) = \int p(y_i^*|\boldsymbol{x}_i;\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}$$

传统模型

1. 构建模型

$\hat{y}_i = f(\boldsymbol{x}_i; \boldsymbol{\theta})$

2. 训练模型

➢ 2.1 样本的损失函数 $\mathrm{loss}(y_i, \hat{y}_i; \boldsymbol{\theta})$

➢ 2.2 数据集的损失函数

$\mathrm{loss}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{loss}(y_i, \hat{y}_i; \boldsymbol{\theta}) + \mathrm{reg}(\boldsymbol{\theta})$

➢ 2.3 训练目标 $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathrm{loss}(\boldsymbol{\theta})$

3. 模型预测

$\hat{y}_i = f(\boldsymbol{x}_i; \hat{\boldsymbol{\theta}})$

概率

贝叶斯模型

1. 构建模型

判别模型 $p(y_i^* | \boldsymbol{x}_i; \boldsymbol{\theta})$

贝叶斯公式

生成模型 $p(y_i^* | \boldsymbol{x}_i; \boldsymbol{\theta}) \propto p(\boldsymbol{x}_i | y_i^*; \boldsymbol{\phi})\, p(y_i^*; \boldsymbol{\psi})$

$p(y_i^* = \text{猫} | \boldsymbol{x}_i) = p(\boldsymbol{x}_i | y_i^* = \text{猫}) * p(y_i^* = \text{猫})$

$p(y_i^* = \text{熊猫} | \boldsymbol{x}_i) = p(\boldsymbol{x}_i | y_i^* = \text{熊猫}) * p(y_i^* = \text{熊猫})$

$p(y_i^* = \text{猫} | \boldsymbol{x}_i) > p(y_i^* = \text{熊猫} | \boldsymbol{x}_i)\ ?\ \hat{y}_i = \text{猫} : \hat{y}_i = \text{熊猫}$

传统模型

1. 构建模型

$$\hat{y}_i = f(\boldsymbol{x}_i; \boldsymbol{\theta})$$

2. 训练模型

- 2.1 样本的损失函数 $\mathrm{loss}(\mathrm{y}_\mathrm{i}, \hat{y}_i; \boldsymbol{\theta})$
- 2.2 数据集的损失函数

$$\mathrm{loss}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{loss}(y_i, \hat{y}_i; \boldsymbol{\theta}) + \mathrm{reg}(\boldsymbol{\theta})$$

- 2.3 训练目标 $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \ \mathrm{loss}(\boldsymbol{\theta})$

3. 模型预测

$$\hat{y}_i = f(\boldsymbol{x}_i; \widehat{\boldsymbol{\theta}})$$

概率

贝叶斯模型

1. 构建模型

判别模型 $p(y_i^* | \boldsymbol{x}_i; \boldsymbol{\theta})$

贝叶斯公式

生成模型 $p(y_i^* | \boldsymbol{x}_i; \boldsymbol{\theta}) \propto p(\boldsymbol{x}_i | y_i^*; \boldsymbol{\phi})\, p(y_i^*; \boldsymbol{\psi})$

2. 训练模型

- 2.2 参数的后验分布

$$p(\theta | \mathcal{D}) \propto p(\mathcal{D} | \theta) * p(\theta)$$

- 2.3 训练目标 $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \ -p(\theta | \mathcal{D})$

3. 模型预测：贝叶斯估计

$$p(y_i^* | \boldsymbol{x}_i; \mathcal{D}) = \int p(y_i^* | \boldsymbol{x}_i; \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathcal{D}) d\boldsymbol{\theta}$$

Thank you !