机器学习第一周作业

68-冯柏淋 201921080330

1. 监督学习有哪些主要类型，并分别举一个例子说明。

Solution. 监督学习主要包括**分类**和**回归**。分类：例如垃圾邮件分类；回归：预测明日气温。

□

2. 请说明在机器学习任务中，训练集，验证集和测试集的作用。

Solution. 训练集用于训练模型，求出模型参数取值；验证集用于确定模型的超参，选出最优模型；而测试集则用于评估最终模型的泛化能力。

□

3. [PRML 1.39] Consider two binary variables x and y having the joint distribution given in Table 1.

		$y=0$	$y=1$
x	0	1/3	1/3
	1	0	1/3

表 1: The joint distribution $p(x, y)$ for two binary variables x and y used in Exercise 1.39.

Evaluate the following quantities.

(a) $\mathrm{H}[x]$ **(c)** $\mathrm{H}[y|x]$ **(e)** $\mathrm{H}[x, y]$

(b) $\mathrm{H}[y]$ **(d)** $\mathrm{H}[x|y]$ **(f)** $\mathrm{I}[x, y]$

Solution.

- $H[x] = -\sum_i p(x_i) \ln p(x_i)$
  (a) $H[x] = -\frac{2}{3} \ln \frac{2}{3} - \frac{1}{3} \ln \frac{1}{3} = 0.802$
  (b) $H[x] = -\frac{1}{3} \ln \frac{1}{3} - \frac{2}{3} \ln \frac{2}{3} = 0.802$
- $H[y|x] = -\sum_x \sum_y p(y, x) \ln p(y|x), \quad H[x, y] = H[y|x] + H[x]$
  (c) $H[y|x] = -(\frac{1}{3} \ln \frac{1}{2} + \frac{1}{3} \ln \frac{1}{2} + \frac{1}{3} \ln 1) = -\frac{2}{3} \ln \frac{1}{2} = 0.462$
  (d) $H[x|y] = H[x, y] - H[y] = 1.264 - 0.802 = 0.462$
  (e) $H[x, y] = H[y|x] + H[x] = 0.462 + 0.802 = 1.264$
- $I[x, y] = -\sum_y \sum_x p(x, y) \ln \frac{p(x) \cdot p(y)}{p(x,y)}$

(f) $I[x,y] = -\frac{4}{3}\ln\frac{2}{3} = 0.541$

Above.

□

4. 什么是共轭先验，它有什么好处，试举出几种分布参数的共轭先验。

Solution. 在贝叶斯概率理论中，如果后验分布与先验分布属于同一个分布族，则先验分布与后验分布称为一对共轭分布，其中的先验分布称为似然函数对应的**共轭先验**。共轭先验分布的提出主要是基于代数计算的便捷考虑，采用共轭先验则可以给出后验分布的闭式解，否则需要通过积分得到后验分布。

例如，伯努利分布的共轭先验分布是 Beta 分布，泊松分布的共轭先验分布是 Gamma 分布；正态分布的共轭先验仍是正态分布等等。

□

5. [PRML 1.11] By setting the derivatives of the log likelihood function Eq.(1) w.r.t. μ and σ^2 equal to zero, verify the results Eq.(2) and Eq.(3).

$$\ln p(\mathbf{x}|\mu,\sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi) \tag{1}$$

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}x_n \tag{2}$$

$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2 \tag{3}$$

Solution.

$$l(\mu,\sigma^2|\mathbf{x}) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

令 $\frac{\partial l(\mu,\sigma^2|x)}{\partial\mu} = 0$，有：

$$-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(2\mu-2x_n) = 0,$$

解得式 (2)，$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}x_n$.

令 $\frac{\partial l(\mu,\sigma^2|x)}{\partial\sigma^2} = 0$，有：

$$\sum_{n=1}^{N}(x_n-\mu)^2\frac{1}{(\sigma^2)^2} - \frac{N}{\sigma^2} = 0,$$

解得式 (3)，$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2$.

□

6. [PRML 1.41] Using the sum and product rules of probability, show that the mutual information $I(x,y)$ satisfies the following relation:

$$I[x,y] = H[x] - H[x|y] = H[y] - H[y|x].$$

Solution.

$$\begin{aligned}
H[y]-H[y|x] &= -\int p(y)\ln p(y)\mathrm{d}y + \iint P(y,x)\ln p(y|x)\mathrm{d}y\mathrm{d}x \\
&= -\left[\int p(y)\ln p(y)\mathrm{d}y + \iint p(y,x)\ln \frac{p(x)}{p(x,y)}\mathrm{d}y\mathrm{d}x\right] \\
&= -\left[\int \left(\int p(x,y)\mathrm{d}x\right)\cdot \ln p(y)\mathrm{d}y + \iint p(y,x)\ln \frac{p(x)}{p(x,y)}\mathrm{d}y\mathrm{d}x\right] \\
&= -\left[\iint p(x,y)\ln p(y)\mathrm{d}x\mathrm{d}y + \iint p(y,x)\ln \frac{p(x)}{p(x,y)}\mathrm{d}y\mathrm{d}x\right] \\
&= -\left[\iint \left(p(x,y)\ln p(y) + p(y,x)\ln \frac{p(x)}{p(x,y)}\right)\mathrm{d}y\mathrm{d}x\right] \\
&= -\iint p(x,y)\ln\left(\frac{p(x)p(y)}{p(x,y)}\right)\mathrm{d}x\mathrm{d}y \\
&= I[x,y].
\end{aligned}$$

$H[x]-H[x|y]=I[x,y]$ 同理可证.

$\square$