HOMEWORK 1

1 证明《统计学习方法》习题 1.2：

通过经验风险最小化推导极大似然估计。证明模型是条件概率分布，当损失函数是对数损失函数时，经验风险最小化等价于极大似然估计。

2 请证明下述 Hoeffding 引理：

Lemma 1. *Let X be a random variable with $E(X) = 0$ and $P(X \in [a, b]) = 1$. Then it holds*

$$E\{\exp(sX)\} \leq \exp\{s^2(b-a)^2/8\}. \tag{0.1}$$

3 已知针对模型 f，使用训练数据集得到的估计记为 $\widehat{f}$，现有独立于训练数据集的 (x_0, y_0)，$y_0 = f(x_0) + \epsilon$，其中 ϵ 为随机误差项，证明：

$$\mathrm{E}\left(y_0 - \widehat{f}\left(x_0\right)\right)^2 = \mathrm{Var}\left(\widehat{f}\left(x_0\right)\right) + \left[\mathrm{Bias}\left(\widehat{f}\left(x_0\right)\right)\right]^2 + \mathrm{Var}(\epsilon).$$

4 Please read the background and then prove the following results.

Background:

Let $\mathbf{y} = \Psi(\mathbf{x})$, where $\mathbf{y}$ is an $m \times 1$ vector, and $\mathbf{x}$ is an $n \times 1$ vector. Denote

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}^\top} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \tag{0.2}$$

Prove the results:

(a) Let $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{y}$ is m×1, $\mathbf{x}$ is n×1, $\mathbf{A}$ is m×n, and $\mathbf{A}$ does not depend on $\mathbf{x}$, then

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}^\top} = \mathbf{A} \tag{0.3}$$

(b) Let the scalar α be defined by $\alpha = \mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, where $\mathbf{y}$ is m×1, $\mathbf{x}$ is n×1, $\mathbf{A}$ is m×n, and $\mathbf{A}$ is independent of $\mathbf{x}$ and $\mathbf{y}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}^{\top}} = \mathbf{x}^{\top}\mathbf{A}^{\top}\left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}^{\top}}\right) + \mathbf{y}^{\top}\mathbf{A} \tag{0.4}$$

(c) For the special case in which the scalar α is given by the quadratic form $\alpha = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ where $\mathbf{x}$ is n×1, $\mathbf{A}$ is n×n, and $\mathbf{A}$ does not depend on $\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}^{\top}} = \mathbf{x}^{\top}(\mathbf{A} + \mathbf{A}^{\top}) \tag{0.5}$$

(d) Let the scalar α be defined by $\alpha = \mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, where $\mathbf{y}$ is $m \times 1$, $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $m \times n$, and both $\mathbf{y}$ and $\mathbf{x}$ are functions of the vector $\mathbf{z}$, where $\mathbf{z}$ is a $q \times 1$ vector and $\mathbf{A}$ does not depend on $\mathbf{z}$. Then

$$\frac{\partial \alpha}{\partial \mathbf{z}^{\top}} = \mathbf{x}^{\top}\mathbf{A}^{\top}\left(\frac{\partial \mathbf{y}}{\partial \mathbf{z}^{\top}}\right) + \mathbf{y}^{\top}\mathbf{A}\left(\frac{\partial \mathbf{x}}{\partial \mathbf{z}^{\top}}\right) \tag{0.6}$$

(e) Let $\mathbf{A}$ be a nonsingular, $m \times m$ matrix whose elements are functions of the scalar parameter α. Then

$$\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \alpha}\mathbf{A}^{-1} \tag{0.7}$$

5 Please write $\hat{\mathbf{a}}$ as the solution of the minimization problem:

$$\min_{\mathbf{a}} \|\mathbf{X}\mathbf{a} - \mathbf{y}\|_2 \tag{0.8}$$

where $\mathbf{X}$ is a $n \times p$ matrix, $\mathbf{y}$ is a $n \times 1$ vector and $\mathbf{a}$ is a $p \times 1$ vector. $\mathbf{X}^{\top}\mathbf{X}$ is nonsingular.

提交时间：9 月 16 日，18:30 之前。请预留一定的时间，迟交作业扣 3 分，作业抄袭 0 分。