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 试利用 Hoeffding 引理证明 Hoeffding 不等式。Hoeffding 引理形式如下：

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 请列举一个实际中有监督学习的应用，请说明（1）问题背景、（2）因变量和自变量分别是什么，以及（3）通过机器学习建模如何解决该实际问题。

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

Background:

Let $\mathbf{y} = \Psi(\mathbf{x})$, where $\mathbf{y}$ is an m-element vector, and $\mathbf{x}$ is an n-element vector. The symbol

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

will denote the m×n matrix of first-order partial derivatives of the transformation from $\mathbf{x}$ to $\mathbf{y}$. Such a matrix is called the Jacobian matrix of the transformation $\Psi()$.

Notice that if $\mathbf{x}$ is actually a scalar then the resulting Jacobian matrix is a m×1 matrix; that is, a single column (a vector). On the other hand, if $\mathbf{y}$ is actually a scalar then the resulting Jacobian matrix is a 1×n matrix; that is, a single row (the transpose of a vector).

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

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

(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

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y}^{\mathrm{T}}\mathbf{A} \tag{0.4}$$

(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

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

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

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \mathbf{y}^{\mathrm{T}}\mathbf{A}\frac{\partial \mathbf{x}}{\partial \mathbf{z}} \tag{0.6}$$

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

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

(4) Please write $\hat{a}$ as the solution of the minimization problem:

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

where $\mathbf{X}$ is a n×p matrix and $\mathbf{y}$ is a n×1 vector. $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is nonsingular.

提交时间：9 月 28 日，晚 20:00 之前。请预留一定的时间，迟交作业扣 3 分，作业抄袭 0 分。