1.《Machine Learning A Probabilistic Perspective》这本书的练习题: Exercise 6.4(10分)

2. Exercise 7.5(20分)

3. Exercise 7.6(20分)

4. Exercise 8.3(15分)

5. Exercise 8.6(25分)

6.把Poisson分布写成指数族的形式，并求解其期望和方差。(10分)

1 Ex6.4 解：

由于 $X \sim \mathcal{N}(\mu, \sigma^2)$，由最大似然的公式：

$$L(\mu,\sigma^2)=\prod_{i=1}^{n} f(x_i;\mu,\sigma^2)=(2\pi\sigma^2)^{-\frac{n}{2}}exp\Big(-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\Big)$$

取对数似然函数，

$$\begin{aligned}\ln L(\mu,\sigma)&=\ln(2\pi\sigma^2)^{-\frac{n}{2}}+\ln exp\Big(-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\Big)\\&=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\end{aligned}$$

对 σ^2 求偏导，

$$\tfrac{\partial \ln L}{\partial \sigma^2}=-\tfrac{n}{2\sigma^2}+\tfrac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0 \qquad \Rightarrow \quad \sigma^2=\tfrac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

由此可以得到，

$$\hat{\sigma}^2_{MLE}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

而

$$E[\hat{\sigma}^2_{MLE}]=E[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2]=\sigma^2$$

故给定 μ，对 σ^2 的最大似然估计是无偏的.

2 Ex7.5 解：

计算负对数似然 NLL：

$$NLL(\mathrm{w},w_0)\propto\sum_{n=1}^{N}(y_n-w_0-\mathrm{w}^Tx_n)^2$$

分别对两个参数求导：

$$\frac{\partial}{\partial w_0}NLL(\mathrm{w},w_0)\propto -Nw_0+\sum_{n=1}^{N}(y_n-\mathrm{w}^Tx_n)$$

$$w_{0,MLE}=\frac{1}{N}\sum_{n=1}^{N}(y_n-\mathrm{w}^Tx_n)=\bar{y}-\mathrm{w}^T\bar{x}$$

对 X 和 y 分别进行中心化处理：

$$X_c=X-\hat{X}$$

$$y_c=y-\hat{y}$$

此时中心化的数据均值都为 0，所以此时线性回归模型中没有 w_0 这一项，同时可得：

$$\mathrm{w}_{MLE}=(X_c^TX_c)^{-1}X_c^Ty_c$$

3 Ex7.6 解：

由第 3 题可知，

$$X_c=X-\hat{X},\quad y_c=y-\hat{y},\quad w_0=\bar{y}-\mathrm{w}^T\bar{x}$$

那么可得

$$w_1=\frac{\sum_i(x_i-\bar{x})(y_i-\bar{y})}{\sum_i(x_i-\bar{x})^2}=\frac{\sum_i x_iy_i-N\bar{x}\bar{y}}{\sum_i x_i^2-N\bar{x}^2}\approx\frac{cov[X,Y]}{var[X]}$$

$$w_0=\bar{y}-w_1\bar{x}\approx E[Y]-w_1E[X]$$

4 Ex8.3 解：

a) $\frac{\partial}{\partial a}\sigma(a) = \frac{exp(-a)}{(1+exp(-a))^2} = \frac{1}{1+e^{-a}}\frac{e^{-a}}{1+e^{-a}} = \sigma(a)(1-\sigma(a))$

b) 由题意，

$$\begin{aligned} g(\mathrm{w}) = \frac{\partial}{\partial \mathrm{w}} NLL(\mathrm{w}) &= \sum_{n=1}^{N} \frac{\partial}{\partial \mathrm{w}}[y_i log\mu_i + (1-y_i)log(1-\mu_i)] \\ &= \sum_{n=1}^{N} y_i \frac{1}{\sigma}\sigma(1-\sigma) - x_i + (1-y_i)\frac{-1}{1-\sigma}\sigma(1-\sigma) - x_i \\ &= \sum_{n=1}^{N} (\sigma(\mathrm{w}^T x_i) - y_i)x_i \end{aligned}$$

c) 对于一个任意的 (形状正确的) 非零向量 u：

$$u^T X^T S X u = (Xu)^T S(Xu)$$

因为 S 为正定矩阵，所以对于任何形状正确的非零 v：

$$v^T S v > 0$$

因为 X 满秩，所以 Xu 非零，所以：

$$(Xu)^T S(Xu) = u^T(X^T S X)u > 0$$

所以 $X^T S X$ 为正定矩阵.

5 Ex8.6 解：

a) F.
$J(\mathrm{w})$ 的第一项的海森矩阵为正定矩阵，第二项的海森矩阵仍为正定矩阵 $(\lambda > 0)$，所以这个函数的海森矩阵正定，故其有全局最优解.

b) F.
考察后验分布的形式：

$$p(\mathrm{w}|D) \propto p(D|\mathrm{w})p(\mathrm{w})$$

$$p(\mathrm{w}) = N(\mathrm{w}|0, \sigma^{-2} I)$$

$$NLL(\mathrm{w}) = -\mathrm{log}p(\mathrm{w}|D) = -\mathrm{log}p(D|\mathrm{w}) + \frac{1}{2\sigma^2}\mathrm{w}^T\mathrm{w} + c$$

所以 $\lambda = \frac{1}{2\sigma^2}$，全局最优解中零的数量和 λ 的取值有关，λ 值负相关于 w 的先验不确定性，其不确定性越小，则 w 越趋向于零向量，所以最后取得的最终解中零也会越多.

c) T.
如果 $\lambda = 0$，即先验不确定性无穷大，后验估计变为最大似然估计。因为此时对于 w 没有限制，所以有可能有 w 的分量趋近于无穷.

d) F.
当 λ 增大时，意味着先验的不确定性减少，所以整个模型的过拟合性质被减弱，一般而言这会导致训练集上的准确率下降。

e) F.
当 λ 增大时，过拟合性质的减弱一般而言能带来在测试集上的准确率上升，不过这并不一定发生。

6 解：

泊松分布的形式为

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

由指数族的定义和标准形式，泊松分布可写成指数族形式

$$p(x|\lambda) = \frac{1}{x!}exp(xlog\lambda - \lambda)$$

对于 X 服从泊松分布，有

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$E(X) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

而

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{(k-1+1)\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda}(\sum_{m=0}^{\infty} \frac{m \cdot \lambda^m}{m!} + \sum_{m=0}^{\infty} \frac{\lambda^m}{m!})(令m = k-1) \\
&= \lambda e^{-\lambda}(\lambda \cdot \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} + \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}) \\
&= \lambda e^{-\lambda}(\lambda e^{\lambda} + e^{\lambda}) \\
&= \lambda(\lambda + 1)
\end{aligned}$$

故有

$$D(X) = E(X^2) - (E(X))^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda$$