P1. Maximum Likelihood Estimation (30 pts: 5+5+10+10)

In the following problems, i.i.d. stands for independent and identically distributed.

(A) Assume that you have a data set $\mathcal{Y}=\{y_i|1\le i\le N\}$ where $y_i\in\{1,2\}$, $i=1,2,\ldots,N$. The data are i.i.d., and follow the following distribution:

$$p(y_i=1)=\theta,\ p(y_i=2)=1-\theta.$$

Suppose that $N=5$, $\mathcal{Y}=\{1,1,2,1,2\}$. **Use MLE to estimate the parameter** θ.

Solution: The likelihood function is $p(\mathcal{Y}|\theta)=\theta^3(1-\theta)^2$, and $\theta_{\mathrm{ML}}=\frac{3}{5}=0.6$. [5pts]

Grading: 3pts for showing that the MLE estimate is the percentage of 1's in the observed data.

(B) Assume that you have a data set $\mathcal{X}=\{x_i|1\le i\le N\}$ where $x_i\in R$, $i=1,2,\ldots,N$. The data are i.i.d., and $x_i\sim\mathcal{N}(\mu,1)$. In other words,

$$p(x_i|\mu)=\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(x-\mu)^2}{2}\}.$$

Suppose that $N=5$, $\mathcal{X}=\{1,2,4,5,6\}$. **Use MLE to estimate the parameter** μ.

Solution: The likelihood function is $p(\mathcal{X}|\theta)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(x_i-\mu)^2}{2}\}$, and

$\mu_{\mathrm{ML}}=\frac{1}{N}\sum_{i=1}^{N}x_i=\frac{18}{5}=3.6$. [5pts]

Grading: 3pts for showing that the MLE estimate is the averaged point.

(C) Assume that we have a data set $\mathcal{D}=\{(x_i,y_i)|1\le i\le N\}$ where each $x_i\in R$ denotes an i.i.d. sample and $y_i\in\{1,2\}$ is the corresponding label. Assume $x_i|y_i=1\sim\mathcal{N}(\mu_1,1)$, $x_i|y_i=2\sim\mathcal{N}(\mu_2,1)$, that is,

$$p(x_i|y_i=1)=\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(x-\mu_1)^2}{2}\},\ p(x_i|y_i=2)=\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(x-\mu_2)^2}{2}\}.$$

In addition, $p(y_i=1)=\theta,\ p(y_i=2)=1-\theta$.

Suppose $N = 6$ and $\mathcal{D} = \{(1,1),(4,1),(5,1),(3,2),(7,2),(8,2)\}$. **Use MLE to estimate the parameter** μ_1, μ_2 and θ.

Solution: The likelihood function is

$$p(\mathcal{D}|\mu_1,\mu_2,\theta) = \prod_{i=1}^{N} p(x_i, y_i|\mu_1,\mu_2,\theta) = \prod_{i=1}^{N} p(y_i|\theta) p(x_i|y_i,\mu_1,\mu_2)$$

$$= \prod_{i=1}^{N} p(y_i|\theta) p(x_i|y_i,\mu_1,\mu_2) = \prod_{i:y_i=1} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(x_i-\mu_1)^2}{2}\}\theta \cdot \prod_{i:y_i=2} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(x_i-\mu_2)^2}{2}\}(1-\theta)$$

Therefore, $\theta = \frac{\sum_{i:y_i=1} 1}{N} = \frac{1}{2}$, [2pts] $\mu_1^{\text{ML}} = \frac{\sum_{i:y_i=1} x_i}{\sum_{i:y_i=1} 1} = \frac{10}{3}$, [4pts] $\mu_2^{\text{ML}} = \frac{\sum_{i:y_i=2} x_i}{\sum_{i:y_i=2} 1} = \frac{18}{3} = 6$. [4pts]

Grading: 1pt for θ for showing that MLE of θ is the percentage of points in class 1. 2pts for each μ_i for showing that MLE of μ_i is the averaged points in class i.

(D) We are drawing i.i.d. data $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ sampled from a uniform distribution $U(-w, w)$, that is,

$$p(x|w) = \begin{cases} \frac{1}{2w} & \text{if } -w \le x \le w \\ 0 & \text{otherwise} \end{cases}$$

Suppose we have observed $\mathcal{X} = \{1,2,3,4\}$. What is the maximum likelihood estimation of the parameter w?

Solution: The likelihood function is $p(\mathcal{X}|w) = \begin{cases} \left(\frac{1}{2w}\right)^4 & \text{if } -w \le x_i \le w \text{ for } i = 1,2,3,4 \\ 0 & \text{otherwise} \end{cases}$.

Since we have observed $\mathcal{X} = \{1,2,3,4\}$, the likelihood is 0 if $w < 4$, and $\left(\frac{1}{2w}\right)^4$ otherwise.

Therefore, $w_{\text{ML}} = 4$. [10pts]

Grading: 3pts for showing that the likelihood is 0 if $w < 4$, and 4pts for MLE of w larger than 4.

P2. Continuous Bayes Classifier (20 pts: 3+3+3+4+4+3)

We want to build a Bayes classifier for a binary classification task ($y=1$ or $y=2$) with a 1-dimensional input feature x. We know the following quantities: (1) $p(y=1)=0.8$; (2) $p(x|y=1)=\frac{1}{3}$ for $2\le x\le 5$ and $p(x|y=1)=0$ otherwise; (3) $p(x|y=2)=\frac{1}{3}$ for $3\le x\le 6$ and $p(x|y=2)=0$ otherwise.

(A) What is the prior for class label $y=2$?

(B) What is $p(y=1|x)$ for $2\le x\le 6$?

(C) What is $p(y=2|x)$ for $2\le x\le 6$?

(D) For $x=2$, what is the class label your classifier will assign? What is the risk of this decision?

(E) For $x=4$, what is the class label your classifier will assign? What is the risk of this decision?

(F) What are the decision regions of your Bayes classifier?

Solution: (A) $p(y=2)=1-p(y=1)=0.2$ [3pts]

(B) When $2\le x<3$,

$$p(y=1|x)=\frac{p(x|y=1)\,p(y=1)}{p(x)}=\frac{p(x|y=1)\,p(y=1)}{p(x|y=1)\,p(y=1)+p(x|y=2)\,p(y=2)}$$

$$=\frac{\frac{1}{3}\times 0.8}{\frac{1}{3}\times 0.8+0\times 0.2}=1$$

When $3\le x\le 5$,

$$p(y=1|x)=\frac{p(x|y=1)\,p(y=1)}{p(x)}=\frac{p(x|y=1)\,p(y=1)}{p(x|y=1)\,p(y=1)+p(x|y=2)\,p(y=2)}$$

$$=\frac{\frac{1}{3}\times 0.8}{\frac{1}{3}\times 0.8+\frac{1}{3}\times 0.2}=\frac{4}{5}$$

When $5 < x \le 6$,

$$p(y=1|x) = \frac{p(x|y=1)\,p(y=1)}{p(x)} = \frac{p(x|y=1)\,p(y=1)}{p(x|y=1)\,p(y=1) + p(x|y=2)\,p(y=2)}$$

$$= \frac{0 \times 0.8}{0 \times 0.8 + \frac{1}{3} \times 0.2} = 0$$

Therefore, $p(y=1|x) = \begin{cases} 1 & 2 \le x < 3 \\ \frac{4}{5} & 3 \le x \le 5 \\ 0 & 5 < x \le 6 \end{cases}$

Grading: 1pt for each case of x. Note that the interval can be either closed or open. i.e. $2 \le x < 3$ could be $2 \le x \le 3$.

(C) By the value of $p(y=1|x)$ and the fact that $p(y=2|x) = 1 - p(y=1|x)$, we have

$$p(y=2|x) = \begin{cases} 0 & 2 \le x < 3 \\ \frac{1}{5} & 3 \le x \le 5 \\ 1 & 5 < x \le 6 \end{cases}$$

Grading: 1pt for each case of x.

(D) By question (B), we know that $p(y=1|x=2) = 1$ and $p(y=2|x=2) = 0$. Since $p(y=1|x=2) > p(y=2|x=2)$, we assign class label $y=1$. [2pts] The risk of decision is $p(y=2|x=2) = 0$. [2pts]

(E) By question (B), we know that $p(y=1|x=4) = \frac{4}{5}$ and $p(y=2|x=4) = \frac{1}{5}$. Since $p(y=1|x=4) > p(y=2|x=4)$, we assign class label $y=1$. [2pts] The risk of decision is $p(y=2|x=4) = \frac{1}{5}$. [2pts]

(F) $\begin{cases} y=1 & 2 \le x \le 5 \\ y=2 & 5 < x \le 6 \\ y = 1 \text{ or } 2 \text{ otherwise} & \end{cases}$

Grading: 1pt for each case of x. For the last case, earn 1pt if the answer is undefined label instead of $y=1 \text{ or } 2$.

P3. Discrete Bayes Classifier (20 pts: 3+3+3+4+4+3)

We want to build a Bayes classifier for a binary classification task ($y=1$ or $y=2$) with one discrete feature x, where $x \in \{0,1,2,3\}$. We know the following quantities: (1) $p(y=1)=0.6$; (2) $p(x=0|y=1)=0.3$, $p(x=1|y=1)=0.1$, $p(x=2|y=1)=0.4$, $p(x=3|y=1)=0.2$; (3) $p(x=0|y=2)=0.4$, $p(x=1|y=2)=0.3$, $p(x=2|y=2)=0.2$, $p(x=3|y=2)=0.1$.

(A) What is the prior for class label $y=2$?

(B) What is $p(y=1|x=3)$?

(C) What is $p(y=2|x=3)$?

(D) For $x=1$, what is the class label your classifier will assign? What is the risk of this decision?

(E) For $x=3$, what is the class label your classifier will assign? What is the risk of this decision?

(F) What are the decision regions of your Bayes classifier?

Solution: (A) $p(y=2)=1-p(y=1)=0.4$ [3pts]

(B)

$$p(y=1|x=3)=\frac{p(x=3|y=1)\,p(y=1)}{p(x=3)}=\frac{p(x=3|y=1)\,p(y=1)}{p(x=3|y=1)\,p(y=1)+p(x=3|y=2)\,p(y=2)}$$

$$=\frac{0.2\times 0.6}{0.2\times 0.6+0.1\times 0.4}=\frac{3}{4}$$

Grading: 3pts for correct answer; earn 2pts if the formula is correct but the answer is wrong.

(C) $p(y=2|x=3)=1-p(y=1|x=3)=\frac{1}{4}$

Grading: 3pts for correct answer; earn 2pts if the formula is correct but the answer is wrong. The formula can also be

$$p(y=2|x=3)=\frac{p(x=3|y=2)\,p(y=2)}{p(x=3|y=1)\,p(y=1)+p(x=3|y=2)\,p(y=2)}$$

(D) Because

$$p(y=1|x=1)=\frac{p(x=1|y=1)\,p(y=1)}{p(x=1)}=\frac{p(x=1|y=1)\,p(y=1)}{p(x=1|y=1)\,p(y=1)+p(x=1|y=2)\,p(y=2)},$$

$$=\frac{0.1\times 0.6}{0.1\times 0.6+0.3\times 0.4}=\frac{1}{3}$$

$p(y=2|x=1)=1-p(y=1|x=1)=\frac{2}{3}$. Since $p(y=2|x=1)>p(y=1|x=1)$, the class label is $y=2$, [2pts] and the risk of this decision is $(y=1|x=1)=\frac{1}{3}$. [2pts]

(E) By question (B), $p(y=1|x=3)=\frac{3}{4}$, $p(y=2|x=3)=1-p(y=1|x=3)=\frac{1}{4}$. Since $p(y=1|x=3)>p(y=2|x=3)$, the class label is $y=1$, [2pts] and the risk of this decision is $p(y=2|x=3)=\frac{1}{4}$. [2pts]

(F) Because

$$p(y=1|x=0)=\frac{p(x=0|y=1)\,p(y=1)}{p(x=0)}=\frac{p(x=0|y=1)\,p(y=1)}{p(x=0|y=1)\,p(y=1)+p(x=0|y=2)\,p(y=2)},$$

$$=\frac{0.3\times 0.6}{0.3\times 0.6+0.4\times 0.4}=\frac{9}{17}$$

the class label for $x=0$ is $y=1$. Also, because

$$p(y=1|x=2)=\frac{p(x=0|y=1)\,p(y=1)}{p(x=2)}=\frac{p(x=2|y=1)\,p(y=1)}{p(x=2|y=1)\,p(y=1)+p(x=2|y=2)\,p(y=2)},$$

$$=\frac{0.4\times 0.6}{0.4\times 0.6+0.2\times 0.4}=\frac{3}{4}$$

the class label for $x = 2$ is $y = 1$. As a result, based on question (D) and (E), the decision regions are

$$\begin{cases} y=1 & x=0 \\ y=2 & x=1 \\ y=1 & x=2 \\ y=1 & x=3 \end{cases}$$

Grading: 1pt for $x = 0$ or $x = 2$, 0.5pt for $x = 1$ or $x = 3$

P4. Naive Bayes Classifier (20 pts: 2+2+2+2+2+2+2+6)

Given the training data in Table 1, we want to train a binary classifier using Naive Bayes, with (1) the last column being the class label y, and (2) each column of X being a binary feature.

Feature $X = (x_1, x_2, x_3)$			Class Label y
Sky	**Humid**	**Wind**	**Enjoy Sport**
sunny	warm	strong	1
rainy	cold	mild	2
sunny	warm	mild	1
rainy	cold	strong	2
sunny	warm	strong	1
rainy	cold	mild	2

Table 1: Training Data Set for Naive Bayes Classifier

(A) What is $p(y = 1)$?

(B) What is $p(x_1 = \text{rainy} | y = 1)$?

(C) What is $p(x_2 = \text{cold} | y = 1)$?

(D) What is $p(x_3 = \text{strong} | y = 1)$?

(E) What is $p(x_1 = \text{rainy} | y = 2)$?

(F) What is $p(x_2 = \text{cold} | y = 2)$?

(G) What is $p(x_3 = \text{strong} | y = 2)$?

(H) Suppose we have a new input vector $x = (\text{sunny}, \text{cold}, \text{strong})$. What is $p(x|y=1) \times p(y=1)$? What is $p(x|y=2) \times p(y=2)$? Which class label will be the Naive Bayes classifier assign to this input?

Solution: (A) $p(y=1) = \frac{3}{6} = \frac{1}{2}$ [2pts]

(B) $p(x_1 = \text{rainy} | y=1) = \frac{0}{3} = 0$ [2pts]

(C) $p(x_2 = \text{cold} | y=1) = \frac{0}{3} = 0$ [2pts]

(D) $p(x_3 = \text{strong} | y=1) = \frac{2}{3}$ [2pts]

(E) $p(x_1 = \text{rainy} | y=2) = \frac{3}{3} = 1$ [2pts]

(F) $p(x_2 = \text{cold} | y=2) = \frac{3}{3} = 1$ [2pts]

(G) $p(x_3 = \text{strong} | y=2) = \frac{1}{3}$ [2pts]

(H) For the new input vector $x = (\text{sunny}, \text{cold}, \text{strong})$,

$$p(x|y=1) \times p(y=1) = p(x_1 = \text{sunny}, x_2 = \text{cold}, x_3 = \text{strong} | y=1) \times p(y=1)$$
$$= p(x_1 = \text{sunny} | y=1)\, p(x_2 = \text{cold} | y=1)\, p(x_3 = \text{strong} | y=1)\, p(y=1)$$

Because $p(x_2 = \text{cold} | y=1) = 0$, $p(x|y=1) \times p(y=1) = 0$. [2pts] Similarly,

$$p(x|y=2) \times p(y=2) = p(x_1 = \text{sunny}, x_2 = \text{cold}, x_3 = \text{strong} | y=2) \times p(y=2)$$
$$= p(x_1 = \text{sunny} | y=2)\, p(x_2 = \text{cold} | y=2)\, p(x_3 = \text{strong} | y=2)\, p(y=2)$$

Because $p(x_1 = \text{rainy} | y=2) = 1$, $p(x_1 = \text{sunny} | y=2) = 0$, so $p(x|y=2) \times p(y=2) = 0$. [2pts]

Because $p(x|y=1) \times p(y=1) = p(x|y=2) \times p(y=2) = 0$, the class label for this input is either $y=1$ or $y=2$. [2pts]

Grading: 2pts if the answer is undefined label instead of either $y=1$ or $y=2$

P5. Optimization (10 pts: 3+3+4)

(A) Let $f(x) = x^2 - 3x + 18$. What is the value of x that solves the following unconstrained optimization problem?

$$\begin{aligned} &\min\ f(x) \\ &s.t. -\infty < x < \infty \end{aligned}$$

(B) Let $f(x) = x^2 - 3x + 18$. What is the value of x that solves the following constrained optimization problem?

$$\begin{aligned} &\min\ f(x) \\ &s.t.\ 4 \le x \le 8 \end{aligned}$$

(C) Let $f(x)$ be a twice continuously differentiable function, and $\bar{x}$ minimizes $f(x)$ in $-\infty < x < \infty$. What is the value of $f'(\bar{x})$, i.e. the derivative of f at $\bar{x}$? Is the second order derivative of f at $\bar{x}$, i.e. $f''(\bar{x})$, negative?

Solution: (A) $f(x) = x^2 - 3x + 18 = \left(x - \frac{3}{2}\right)^2 + \frac{63}{4}$. Because $\left(x - \frac{3}{2}\right)^2 \ge 0$, $f(x) \ge \frac{63}{4}$, and the value of x that solves the optimization problem is $x^{\text{opt}} = \frac{3}{2}$. [3pts]

Grading: 1pt for knowing the correct formula, e.g. setting the derivative to zero, but the final answer is wrong.

(B) $f(x) = x^2 - 3x + 18 = \left(x - \frac{3}{2}\right)^2 + \frac{63}{4}$, so $f(x)$ is monotonically increasing when $4 \le x \le 8$, and the value of x that solves the optimization problem is $x^{\text{opt}} = 4$. [3pts]

Grading: 1pt for reasonable argument while the final answer is wrong.

(C) $f'(\bar{x}) = 0$. [2pts] $f''(\bar{x})$ cannot be negative, because $f''(\bar{x}) \ge 0$. [2pts]

Grading: While this problem is not supposed to be solved by assuming $f(x) = x^2 - 3x + 18$ ($f(x)$ is a general differential function here), students still earn full points if they use $f(x) = x^2 - 3x + 18$ to give correct answers.