Problem 0 [1 bonus mark]. Write your name and student ID (number) on every submitted answer sheet.

Problem 1 [10+10=20 marks]. Let x be a discrete variable.

a) Give the formula for $\mathbb{E}_{x\sim P(x)}[f(x)]$. No proof is needed.

$$\mathbb{E}_{x\sim p(x)}[f(x)] = \sum_x p(x) f(x)$$

b) Consider two distributions $P_1(x)$ and $P_2(x)$. We define a new distribution $P(x) = \lambda P_1(x) + (1-\lambda)P_2(x)$ for $\lambda \in [0,1]$.

Prove that
$\mathbb{E}_{x\sim P(x)}[f(x)] = \lambda \mathbb{E}_{x\sim P_1(x)}[f(x)] + (1-\lambda)\mathbb{E}_{x\sim P_2(x)}[f(x)]$

$$\text{LHS} = \mathbb{E}_{x\sim p(x)}[f(x)]$$

$$= \sum_x p(x) f(x)$$

$$= \sum_x [\lambda P_1(x) + (1-\lambda) P_2(x)] f(x)$$

$$= \sum_x [\lambda P_1(x) f(x) + (1-\lambda) P_2(x) \cdot f(x)]$$

$$= \sum_x \lambda P_1(x) f(x) + \sum_x (1-\lambda) P_2(x) f(x)$$

$$= \lambda \mathbb{E}_{x\sim P_1(x)}[f(x)] + (1-\lambda) \mathbb{E}_{x\sim P_2(x)}[f(x)]$$

Problem 2. [10+20=30 marks] Consider a binary random variable $x \in \{0,1\}$. A Bernoulli distribution is characterized by a scalar parameter $\pi \in [0,1]$. The probability distribution of x is given by $P(x=1) = \pi$ and $P(x=0) = 1-\pi$. The two cases can be unified as $P(x) = \pi^x(1-\pi)^{1-x}$.

a) Derive the likelihood of π on the dataset $\mathcal{D} = \{x^{(m)}\}_{m=1}^M$, where samples are independent and identically distributed (iid).

$$\mathcal{L}(\pi; \mathcal{D}) = p(\mathcal{D}; \pi) = \prod_{m=1}^{M} P(x^{(m)}; \pi)$$

$$= \prod_{m=1}^{M} \pi^{x^{(m)}} (1-\pi)^{1-x^{(m)}}$$ [ok]

$$= \pi^{\sum_{m=1}^M x^{(m)}} (1-\pi)^{M - \sum_{m=1}^M x^{(m)}}$$

b) Derive the closed-form solution of the maximum likelihood estimation of π on $\mathcal{D}$.

$$\log \mathcal{L}(\pi; \mathcal{D}) = M_1 \log \pi + (M - M_1) \log(1-\pi)$$

$$\frac{\partial}{\partial \pi} \log \mathcal{L}(\pi; \mathcal{D}) = M_1 \cdot \frac{1}{\pi} + (M - M_1) \cdot \frac{1}{1-\pi} \cdot (-1) \overset{\text{set}}{=} 0$$

$$M_1 - \cancel{M_1 \pi} - M\pi + \cancel{M_1 \pi} = 0$$

$$\pi = \frac{M_1}{M}$$

where M_1 is defined as $M_1 = \sum_{m=1}^{M} x^{(m)}$

Problem 3 [10+10+10+20=50 marks].

a) Give the formal definition of a convex set. *Hint:* Intuitively, a convex set means that, for every two points in the set, any middle point is also in the set.

S is a convex set if

for every x, y in S, for every $\lambda \in (0,1)$

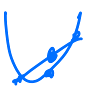
$\lambda x + (1-\lambda) y$ is also in S.

b) Give the formal definition of a convex function. *Hint*: Intuitively, a convex function means that the average of function values is less than or equal to the function value of the average input.

f is a convex function if

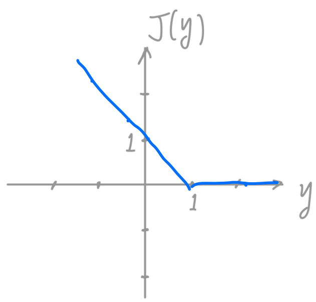
① domain of f is a convex set

② for every $x, y \in \mathrm{dom} f$, for every $\lambda \in (0,1)$



$$\lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda) y)$$

c) A hinge loss function is commonly used in machine learning. Consider $J(y) = \max\{0, 1-y\}$ for $y \in \mathbb{R}$, where $\max\{a, b\}$ chooses the maximum value between a and b. Draw the function $J(y)$ in the right plot.



$\max\{0, 1 - (\lambda y_1 + (1-\lambda) y_2)\}$

d) Prove that $J(y)$ is a convex function in y. **Requirement:** Rigorous derivations are needed.

$\mathrm{dom}\, J = \mathbb{R}$ convex. (may be omitted)

Consider any two points $y_1, y_2 \in \mathbb{R}$ and $\lambda \in (0,1)$

$\lambda J(y_1) + (1-\lambda) J(y_2)$

$= \max\{0, \lambda - \lambda y_1\} + \max\{0, (1-\lambda) - (1-\lambda) y_2\}$ (1)

$J(\lambda x + (1-\lambda) y_2) = \max\{0, 1 - [\lambda y_1 + (1-\lambda) y_2]\}$ (2)

Goal is to show (1) $\geq$ (2)

We see Eq. (1) $\geq 0 + 0 = 0$. (3)

Eq. (1) $\geq \lambda - \lambda y_1 + (1-\lambda) - (1-\lambda) y_2$

$= 1 - [\lambda y_1 + (1-\lambda) y_2]$ (4)

Combing (3) and (4), we can conclude

(1) $\geq$ (2)

END OF THE EXAM

Scrap paper

- Can be detached
- Can be used as an answer sheet. If so, please
  - Mark the problem ID clearly,
  - Write your name on every submitted answer sheet, and
  - Ask TAs to staple all sheets by the end of the exam