CSE 251A Winter 2021 - Take Home Exam 4

1 hour **Name:** **PID:**

---

(1) This is an open book, open notes exam. You are free to consult any text book or notes. **You are not allowed to consult with any other person.**

(2) If you need any clarification, please post a private message to the instructors on Piazza.

(3) Remember that your work is graded on the *clarity* of your writing and explanation as well as the validity of what you write.

(4) This is a one-hour exam.

(1) A decision region of a classifier f corresponding to a label y is the set of inputs where f predicts y. Formally, the decision region for a classifier $f : \mathbb{R}^d \to \mathcal{Y}$ for a label y is the set $\{x \in \mathbb{R}^d, f(x) = y\}$.

We say that a classifier has convex decision regions if all its decision regions corresponding to every label are convex.

(a) (5 points) Suppose we are given a multiclass linear classifier f_L with $k > 2$ classes. Does f_L have convex decision regions? Please state which of these choices is true: 1) true for all f_L; 2) true for some f_L but not others; 3) not true for any f_L. Justify your answer.

True for all f_L. Let $\mathcal{Y} = \{1, \cdots, k\}$. For any $y \in \mathcal{Y}$, f_L predicts an $x \in \mathbb{R}^d$ as y when

$$w_y^\top x + b_y > w_i^\top x + b_i, \forall i \neq y.$$

Now, suppose f_L predicts both x_1, x_2 as y. Let $x' = tx_1 + (1-t)x_2$ for $t \in (0,1)$. Then, for any $i \neq y$,

$$\begin{aligned} w_y^\top x' + b_y &= t(w_y^\top x_1 + b_y) + (1-t)(w_y^\top x_2 + b_y) \\ &> t(w_i^\top x_1 + b_i) + (1-t)(w_i^\top x_2 + b_i) \\ &= w_i^\top x' + b_i. \end{aligned}$$

Therefore, f_L predicts x' as y. This shows that f_L have convex decision regions.

(b) (5 points) Now suppose f_T is a decision tree with two output labels. Does f_T have convex decision regions? Please state which of these choices is true: 1) true for all f_L; 2) true for some f_L but not others; 3) not true for any f_L. Justify your answer.

True for some f_L but not others. Consider $\mathbb{R}^2$. A decision tree that predicts 0 when $x < 0$ and predicts 1 when $x \geq 0$ has convex decision regions. However, a decision tree that predicts 0 when $x < 0$ and $y < 0$ and predicts 1 when $x \geq 0$ or $y \geq 0$ does not have convex decision regions.

(2) State whether the following statements are true or false. Justify your answer.

(a) (5 points) If $K(x,z)$ is a kernel, then $L(x,z) = K(x,x) + K(z,z) - 2K(x,z)$ is also a kernel.

False. Let $K(x,z) = \phi(x)^\top \phi(z)$. Let $\phi(x) \neq \phi(z)$. Then, we have

$$L(x,z) = L(z,x) = (\phi(x) - \phi(z))^\top (\phi(x) - \phi(z)) > 0,$$

and

$$L(x,x) = L(z,z) = 0.$$

Then, the matrix

$$A = \begin{pmatrix} L(x,x) & L(x,z) \\ L(z,x) & L(z,z) \end{pmatrix}$$

is not PSD. This is because if we choose $v = (1,-1)^\top$, $v^\top A v = -2L(z,x) < 0$.

(b) (5 points) Suppose $K(x,z)$ is a kernel. Then $K(x,x)$ is always a convex function of x.

False. Consider $x \in \mathbb{R}$. Let $K(x,z) = \phi(x) \cdot \phi(z)$ where $\phi : \mathbb{R} \to \mathbb{R}$. Then, $K(x,x) = \phi(x)^2$. Let $h(x)$ be a non-negative non-convex function (such as the standard Gaussian pdf) and $\phi(x) = \sqrt{h(x)}$. Then, $K(x,x)$ is not a convex function of x.