**Bingjun Guo (bingjun3)**

# 0 Instructions

Homework is due Thursday, February 6, 2024 at 23:59pm Central Time. Please refer to `https://courses.grainger.illinois.edu/cs446/sp2024/homework/hw/index.html` for course policy on homeworks and submission instructions.

# 1 Short answer: 10pts

1. $O(MNd)$
2. $k = 10$
3. $\left(\begin{bmatrix}1\\1\end{bmatrix}, 1\right)$
4. The largest eigenvalue of $A^\top A$ is the square of the largest singular value of $A$.
5. In sentiment recognization tasks on natural language, for example, movie comments, for the probability of the phrase "not good" appearing in a positive comment:

$$P(\text{"}not\text{"}, \text{"}good\text{"}|positive) \neq P(\text{"}not\text{"}|positive) \cdot P(\text{"}good\text{"}|positive)$$

since both $P(\text{"}not\text{"}|positive)$ and $P(\text{"}good\text{"}|negative)$ are adequately high but "not good" should appear really rare in positive comments.

# 2 Linear Regression: 10pts

1. $X$ can be considered as a linear transform from $\mathbb{R}^n$ to $\mathbb{R}^d$. Thus, it complies to the Rank-Nullity Theorem:

$$\text{rank}(X) + \text{nullity}(X) = n$$

in which the dimention of the input space is $n$ and $\text{rank}(X) = n$.
Therefore, $\text{nullity}(X) = 0$, which indicates that $X$ is invertible. Thus, there exists $\boldsymbol{w} = X^{-1}\boldsymbol{y}$ such that satisfies $X\boldsymbol{w} = \boldsymbol{y}$.

2. Since the number of non-zero singular values of $A$ equals to $\text{rank}(A)$, $\Sigma$ is a diagonal matrix consists of positive singular values of $A$, and $X$ is real, $\text{rank}(\Sigma) = \text{rank}(A) = n$.
3. Firstly we will prove that $X^\top$ and $XX^\top$ share the same nullity,
*i.e.*, $X^\top M = 0 \iff XX^\top M = 0$ for $M \in \mathbb{R}^n$.

$X \cdot 0 = 0$, thus $X^\top M = 0 \to X(X^\top M) = 0 \to XX^\top M = 0$.
Supoose $XX^\top M = 0$, then we have $M^\top XX^\top M = 0$, and thus $(X^\top M)^\top X^\top M = 0$, with $X^\top M \in \mathbb{R}^d$. $(X^\top M)^\top X^\top M$ equals to sum of square of all the entries in $X^\top M$, which can only be greater or equal to 0 since its a real vector. Thus, all entries in $X^\top M$ are 0, *i.e.*, $X^\top M = \mathbf{0}$. Therefore, $XX^\top M = 0 \to X^\top M = 0$.
Secondly, since nullity$(X^\top)$ = nullity$(XX^\top)$, according to the Rank-Nullity Theorem introduced in the first question, since the dimention of input space (RHS of the equation) is both $n$ for linear transforms $X^\top : \mathbb{R}^n \to \mathbb{R}^d$ and $XX^\top : \mathbb{R}^n \to \mathbb{R}^n$, rank$(XX^\top)$ = rank$(X^\top)$ = rank$(X) = n$. Thus, $XX^\top$ is a full-rank square matrix, *i.e.*, $XX^\top$ is invertible.

# 3 SVM: 10 pts

1. 2, which happens in the case that the closest two vectorsa in different class are selected and there's only such 2 points in $\mathcal{D}$ belonging to different classes that have such distance between each other.

2. The largest possible **TO BE DONE**

**3. (a)** $\phi(\boldsymbol{x}) = (x_1^2, x_2^2, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1)$
**(b)**
$\phi(-1,-1) = (1, 1, \sqrt{2}, -\sqrt{2}, -\sqrt{2}, 1)$
$\phi(1,1) = (1, 1, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}, 1)$
$\phi(1,-1) = (1, 1, -\sqrt{2}, \sqrt{2}, -\sqrt{2}, 1)$
$\phi(-1,1) = (1, 1, -\sqrt{2}, -\sqrt{2}, \sqrt{2}, 1)$
**Therefore, $\boldsymbol{w}$ can be $(0, 0, 1, 0, 0, 0)$.**

# 4 Gaussian Naive Bayes: 15pts

1.
$$\frac{1}{1 + \exp(\log \frac{A}{B})} = \frac{B}{B + A}$$

$$P(y = +1|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|y = +1) \cdot p}{P(\boldsymbol{x})} = \frac{P(\boldsymbol{x}|y = +1) \cdot p}{P(\boldsymbol{x}|y = +1) \cdot p + P(\boldsymbol{x}|y = -1) \cdot (1 - p)}$$

Therefore, for $B = P(\boldsymbol{x}|y = +1) \cdot p$ and $A = P(\boldsymbol{x}|y = -1) \cdot (1 - p)$,

$$P(y = +1|\boldsymbol{x}) = \frac{1}{1 + \exp(\log \frac{A}{B})}$$

2. **TO BE DONE**

**3.**

$$P(y|\boldsymbol{x}) = \frac{1}{1 + \mathbf{exp}(y \cdot (\boldsymbol{w}^\top \boldsymbol{x} + b))}$$

# 5 Linear regression: 14pts + 1pt