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. $O(MN)$. Each N test image has to be compared to all M training images to get the top l nearest neighbors.
2. 10. Since if the number is too small, the predicted label will be too easily to change.
3. $\omega = (1,1)^T$, $b = 0$.
4. The largest eigenvalue of A^TA is the square of the largest singular value of A.
5. Text data. Since real world text data often involves sequential relationships between words and paragraphs, which violates the iid assumption.

2 Linear Regression: 10pts

1. Since $\omega \in \mathbb{R}^d$, thus the number of unkown parameters is d. While the rank of the matrix X is n, thus the number of valid equations is n. Since $n < d$, there are infinite solutions.
2. From the definition of SVD, we know the rank of $\sum$ is the same as $rank(X) = n$.
3. From the SVD result, we have $XX^T = U\Sigma V^T V\Sigma U^T$, since V is an orthogonal matrix, thus $V^TV = I$. Then we have $XX^T = U\Sigma\Sigma^T U^T$. We can view this as an SVD result of XX^T, thus the rank of XX^T is n, which means XX^T is full rank.

3 SVM: 10 pts

1. 2. Since the dataset is linearly separable, there must be at least two support vectors that each of them is on the margin.
2. The smallest possible number would be 3 and the largest is n. Since all positive αs must be on the margin, so the smallest number is 3. If all data points are on the margin, but their contributions to this solution is 0, then the largest number is n.

3. (a)

$$\begin{aligned}(x^Tz+1)^T(x^Tz+1) &= (z^Tx+1)(x^Tz+1)\\ &= z^Txx^Tz+z^Tx+x^Tz+1\\ &= (x^Tz)^Tx^Tz+(\sqrt{2}x)^T(\sqrt{2}z)+1\\ &= (x_1^2,x_2^2,\sqrt{2}x_1,\sqrt{2}x_2,1)^T(z_1^2,z_2^2,\sqrt{2}z_1,\sqrt{2}z_2,1)\end{aligned}$$

Thus, the feature map is $\phi(x)=(x_1^2,x_2^2,\sqrt{2}x_1x_2,\sqrt{2}x_1,\sqrt{2}x_2,1)$

(b) One solution is $\omega=(0,0,1,0,0,0)$

4 Gaussian Naive Bayes: 15pts

1. From naive bayes equation, we have:

$$\begin{aligned}P(y=1|x) &= \frac{P(x|y=1)P(y=1)}{P(x)}\\ &= \frac{P(x|y=1)p}{P(x|y=1)p+P(x|y=-1)(1-p)}\\ &= \frac{1}{1+\frac{P(x|y=-1)(1-p)}{P(x|y=1)p}}\end{aligned}$$

Thus, $A=P(x|y=-1)(1-p)$, $B=P(x|y=1)p$.

2. From the results of the previous question, we have:

$$\begin{aligned}\log\frac{A}{B} &= \log\frac{(1-p)\prod_j e^{-\frac{1}{2}(x_j-\mu_{-,j})^2}}{p\prod_j e^{-\frac{1}{2}(x_j-\mu_{+,j})^2}}\\ &= \log\frac{1-p}{p}+\sum_j\left(-\frac{1}{2}(x_j-\mu_{-,j})^2+\frac{1}{2}(x_j-\mu_{+,j})^2\right)\\ &= \log\frac{1-p}{p}+\sum_j\frac{1}{2}\left(x_j^2-2x_j\mu_{+,j}+\mu_{+,j}^2-x_j^2+2x_j\mu_{-,j}-\mu_{-,j}^2\right)\\ &= \log\frac{1-p}{p}+\frac{1}{2}(\mu_--\mu_+)^Tx+\frac{1}{2}({\mu_+}^T\mu_+-{\mu_-}^T\mu_-)\end{aligned}$$

3. From the previous results, we can easily know a similar result for $P(y=-1|x)$:

$$\begin{aligned} P(y=-1|x) &= \frac{1}{1+e^{\log \frac{B}{A}}} \\ &= \frac{1}{1+e^{-\log \frac{A}{B}}} \\ &= \frac{1}{1+e^{-(\omega^T x+b)}} \end{aligned}$$

Combined with $P(y=1|x) = \frac{1}{1+e^{\omega^T x+b}}$, we get the single expression:

$$P(y|x) = \frac{1}{1+e^{-y(\omega^T x+b)}} \tag{1}$$

where ω and b are as same as the previous question.

5 Linear regression: 14pts + 1pt

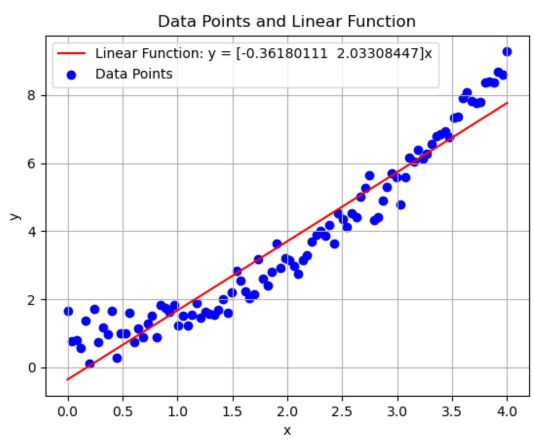

Figure 1: Results of the gradient descent method

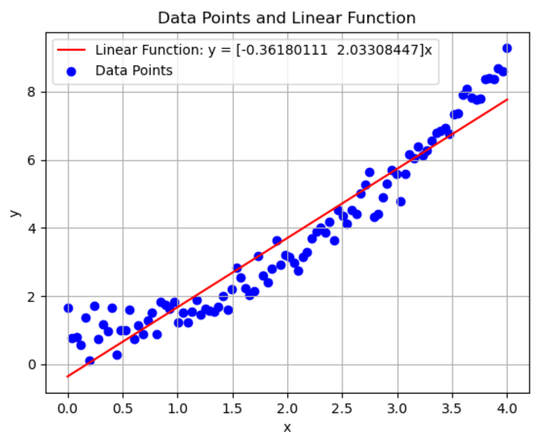

Figure 2: Results of the gradient descent method