Homework 2

Yixuan Li
CSE 151A: Machine Learning and Algorithms

May 14, 2025

1 Conceptual and Mathematical Problems

1. (a) $||x||_1 = 1 + 2 + 3 + 4 = 10$
(b) $||x||_2 = \sqrt{1^2 + 2^2 + 3^2 + 4^2} = \sqrt{30}$
(c) $||x||_\infty = \max_i |x_i| = 4$

2. *Comparing the norms.*
(a) The point $x = [x_1.x_2, ..., x_d]^T$ with $|x_1| = |x_2| = ... = |x_d| = 1$, has the largest l_1 norm and l_2 norm:

- $l_1 = \sum_{i=1}^{d} |x_i| = d$
- $l_2 = \sqrt{\sum_{i=1}^{d} x_i^2} = \sqrt{d}$

(b) The point $x = [x_1.x_2, ..., x_d]^T$ with $|x_1| = |x_2| = ... = |x_d| = \frac{1}{\sqrt{d}}$, has the largest l_1 norm:

- $l_1 = \sum_{i=1}^{d} |x_i| = \sqrt{d}$

The point $x = [x_1.x_2, ..., x_d]^T$ with one of the $|x_i| = 1$, while other $|x_i| = 0$, has the largest l_∞ norm:

- $l_\infty = \max_i |x_i| = 1$

3. It is not a metric.
It violates the fourth rule: the triangle inequality.
For example, $d(B, C) = 4,\ d(B, A) = 2,\ d(A, C) = 1$, which does not satisfy the triangle inequality: $d(B, C) \leq d(B, A) + d(A, C)$.

4. (a) No, it is not a metric. It violates the first and the third rule, while $d(x, y) = x - y$ can be smaller than zero if $x < y$, and $d(x, y) = x - y$ does not equal $d(y, x) = y - x$.
(b) Yes, it is a metric.
(c) No, it is not a metric. It violates the fourth rule: the triangle inequality. For example, consider the case m=1, let x=1, y=3, z=5, we have $d(x, z) = (1 - 5)^2 = 16,\ d(x, y) =$

$(1-3)^2 = 4,\ d(y,z) = (3-5)^2 = 4$, and it is clear that $d(x,z) > d(x,y) + d(x,z)$.

5. KL divergence

$$D_{\mathrm{KL}}(p,q) = \sum_i P(i)\log\frac{P(i)}{Q(i)} = \frac{1}{2}\times log2 + \frac{1}{4}\times 0 + \frac{1}{8}\times log\frac{3}{4} + \frac{1}{16}\times log\frac{3}{8} + \frac{1}{16}\times log\frac{3}{8} = \frac{1}{4}log3 - \frac{1}{8}log2$$

6. (a) Classification. The reason is that the prediction metric is the accuracy of judging the person's action.

(b) Regression. The reason is that the prediction metric is how close the estimated speed value is to the true speed of the car.

(c) Regression. The reason is that the prediction metric is how close the estimated GPA is to the actual GPA during the freshman year of college.

(d) Classification. The reason is that the output of the model is the label "Complete" or "Not Complete", instead of a number or a value.

7. (a)

$$E(x) = -1\times\frac{1}{3} + 0 + 1\times\frac{1}{3} = 0,\ E(y) = -1\times\frac{1}{3} + 0 + 1\times\frac{1}{3} = 0$$

$$E(xy) = \sum_{x,y} xyf(x,y) = -1\times\frac{1}{3} - 1\times\frac{1}{3} + 0 = -\frac{2}{3}$$

$$Cov(x,y) = E(xy) - E(x)E(y) = -\frac{2}{3}$$

(b)

$$\sigma_x = \sigma_y = \sqrt{\frac{1+1}{3}} = \sqrt{\frac{2}{3}}$$

$$\rho = \frac{Cov(x,y)}{\sigma_x.\sigma_y} = -1$$

8. There are two possible reasons:

1. There might be much more "+" data than "-" data in the training data set, like 95% is "+" and 5% is "-". Since we classify based on $\pi_i P(x_i)$, while $\pi_+ \gg \pi_-$, we might predict "+" at all points. (This is the main reason)

2. There may be not enough features used to predict the result. For example, we only use one feature to predict and the distribution of the train data in two classes might be fully overlapped. (But this is not the main reason)

9. *Wine classification.*
   (a) Class 2, because $\max_i \pi_i P_i(x) = \pi_2 P_2(x) \approx 0.6 \times 0.39$.
   (b) Class 2, because $\max_i \pi_i P_i(x) = \pi_2 P_2(x) \approx 0.6 \times 0.39$.
   (c) Class 3, because $\max_i \pi_i P_i(x) = \pi_3 P_3(x) \approx 0.7 \times 0.28$.
   (d) Class 1, because $\max_i \pi_i P_i(x) = \pi_1 P_1(x) \approx 0.7 \times 0.33$.
   (e) Class 1, because $\max_i \pi_i P_i(x) = \pi_1 P_1(x) \approx 0.8 \times 0.33$.

2 Programming Problems

10. *Cross-validation for nearest neighbor classification.*
   (a) The accuracy of the classifier is: 76.97%.

The 3×3 confusion matrix is:

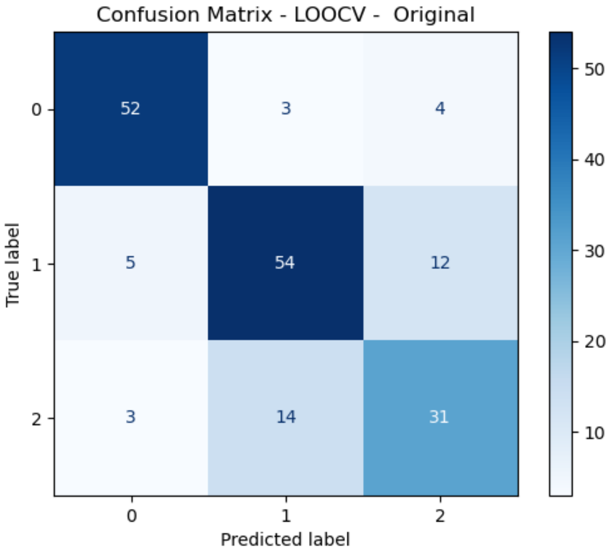

Figure 1: Problem 10 (a) Confusion matrix

(b) The plot of the estimates is as follows:

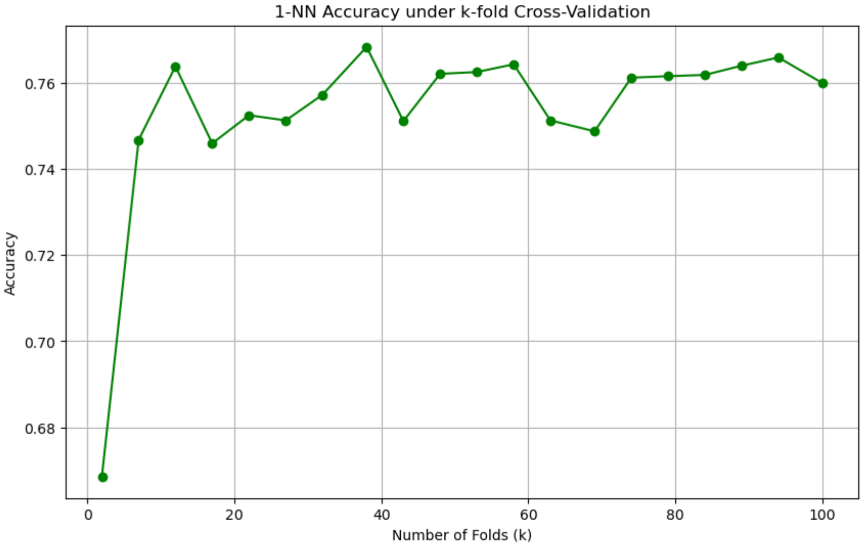

Figure 2: Problem 10 (b) Plot of the estimates

(c) The accuracy of the classifier after feature normalization is: 94.94%.

The confusion matrix after normalization is:

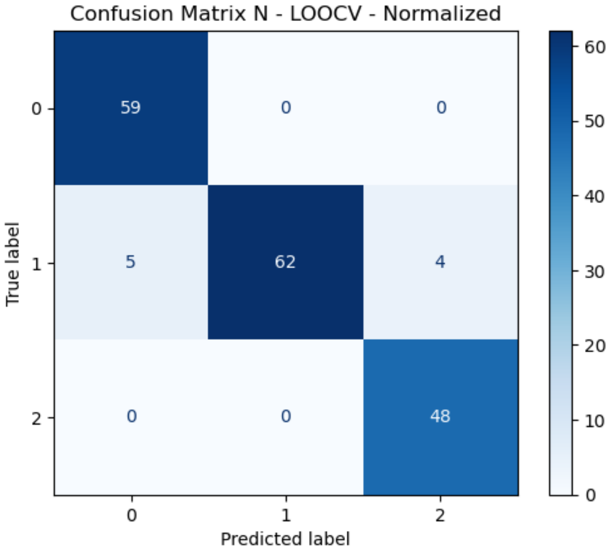

Figure 3: Problem 10 (a) Confusion matrix after normalization