Homework 4

Yixuan Li
CSE 151A: Machine Learning and Algorithms

May 14, 2025

1 Conceptual and Mathematical Problems

1. *Identical spherical Gaussian*

Classify Class 1:

$$\pi_1 p_1(x) \geq \pi_2 p_2(x)$$
$$\log \pi_1 p_1(x) \geq \log \pi_2 p_2(x)$$
$$\log p_1(x) - \log p_2(x) \geq \log \pi_2 - \log \pi_1$$

Since it is spherical Gaussian: $\Sigma_1 = \Sigma_2 = \Sigma$,
then

$$\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) - \frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) \geq \log\frac{\pi_2}{\pi_1}$$

expand:

$$\mu_1^T\Sigma^{-1}x - \mu_2^T\Sigma^{-1}x \geq \log\frac{\pi_2}{\pi_1} + \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 - \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2$$

group terms:

$$\left(\Sigma^{-1}(\mu_1-\mu_2)\right)^T x \geq \log\frac{\pi_2}{\pi_1} + \frac{1}{2}\left(\mu_1^T\Sigma^{-1}\mu_1 - \mu_2^T\Sigma^{-1}\mu_2\right)$$

Define:

$$w = \Sigma^{-1}(\mu_1-\mu_2)$$
$$b = \log\frac{\pi_2}{\pi_1} + \frac{1}{2}\left(\mu_1^T\Sigma^{-1}\mu_1 - \mu_2^T\Sigma^{-1}\mu_2\right)$$

Next, for spherical Σ,

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{d\times d}$$

thus

$$\Sigma^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{d\times d} = \frac{1}{\sigma^2}I_d$$

Thus:

$$w = \Sigma^{-1}(\mu_1 - \mu_2) = \frac{1}{\sigma^2}(\mu_1 - \mu_2)$$

$$b = \log \frac{\pi_2}{\pi_1} + \frac{1}{2\sigma^2}\left(\|\mu_1\|^2 - \|\mu_2\|^2\right)$$

2. *Example of regression with one predictor variable.*

(a) To predict y without knowledge of x, we will predict $y = y_{mean} = \frac{1+3+4+6}{4} = 3.5$

$$MSE = \frac{(1-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (6-3.5)^2}{4} = 3.25$$

(b) To predict y based on x with the regression function $y = x$, the predicted $y_1 = 1,\ y_2 = 1,\ y_3 = 4,\ y_4 = 4$.

$$MSE = \frac{(0)^2 + (2)^2 + (0)^2 + (2)^2}{4} = 2$$

(c) Minimize the MSE (equal to minimizing the square loss):

$$\min_{a,b} \quad L(a,b) = \sum_{i=1}^{4} \left(y^{(i)} - (a \cdot x^{(i)} + b)\right)^2$$

$$= (1-a-b)^2 + (3-a-b)^2 + (4-4a-b)^2 + (6-4a-b)^2$$

Take derivatives and set to zero:

$$\text{Let} \quad \frac{\partial L}{\partial a} = 0, \quad \frac{\partial L}{\partial b} = 0$$

Compute derivatives:

$$\frac{\partial L}{\partial a} = -2(1-a-b) - 2(3-a-b) - 8(4-4a-b) - 8(6-4a-b)$$

$$= 68a + 20b - 88 = 0 \quad (1)$$

$$\frac{\partial L}{\partial b} = -2(1-a-b) - 2(3-a-b) - 2(4-4a-b) - 2(6-4a-b)$$

$$= 20a + 8b - 28 = 0 \quad (2)$$

Solve (1) and (2) together:

$$a = 1, \quad b = 1$$

Thus:

$$y = x + 1$$

Predictions:

$$y_1 = 2, \quad y_2 = 2, \quad y_3 = 5, \quad y_4 = 5$$

Compute Mean Squared Error (MSE):

$$\text{MSE} = \frac{(2-1)^2 + (3-2)^2 + (5-4)^2 + (5-6)^2}{4} = 1$$

3. *Optimality of the mean.*

(a)

$$L(S) = \frac{1}{n}\sum_{i=1}^{n}(x_i - s)^2$$

$$\frac{\partial L}{\partial s} = -\frac{2}{n}\sum_{i=1}^{n}(x_i - s) = -\frac{2\sum_{i=1}^{n} x_i}{n} + 2s = -2\bar{x} + 2s$$

(b)

$$\text{Let} \quad \frac{\partial L}{\partial s} = 0,$$
$$s = \bar{x}$$

4.

$$L(w, b) = |y - \hat{y}| = \sum_{i=1}^{n} |y^{(i)} - (wx^{(i)} + b)|$$

5. *Writing expression in matrix-vector form.*

(a)

$$\frac{y^{(1)} + ... + y^{(n)}}{n} = \frac{y^T\mathbf{1}}{n} = \frac{\mathbf{1}^T y}{n}$$

(b)

$$XX^T$$

(c)

$$\frac{x^{(1)} + ... + x^{(n)}}{n} = \frac{x^T\mathbf{1}}{n} = \frac{\mathbf{1}^T x}{n}$$

(d)

$$\frac{X^T X}{n}$$

2 Programming Problems

6. *Optimality of the mean.*

(1) A linear boundary that is not parallel to the coordinate axes is shown as follows, with $\mu_1 = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$, $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} -5 \\ 1 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

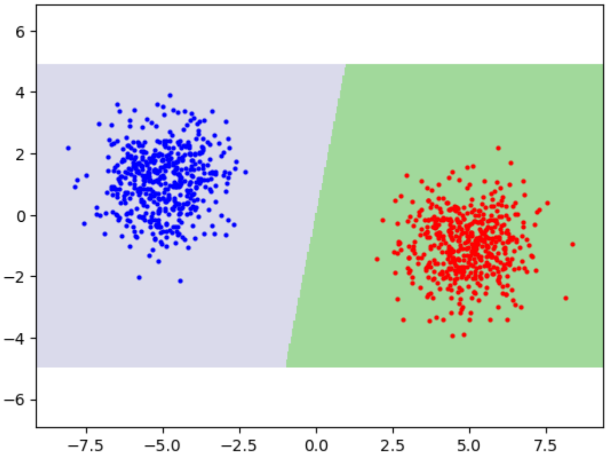


Figure 1: Problem 6(a). A linear boundary that is not parallel to the coordinate axes

(2) A spherical boundary is shown as follows, with $\mu_1 = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$, $\Sigma_1 = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

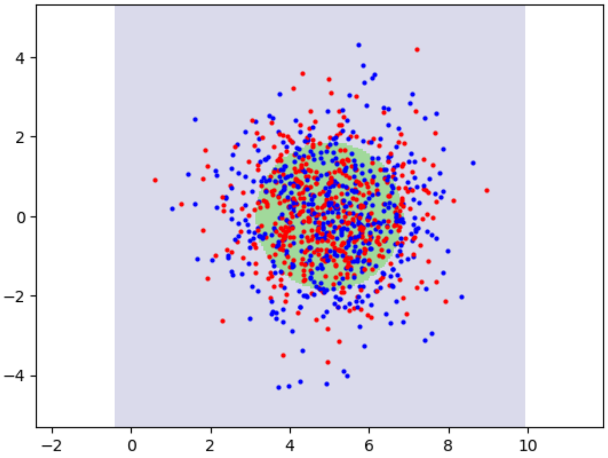


Figure 2: Problem 6(b). A spherical boundary

(3) A boundary that is elliptical is shown as follows, with $\mu_1 = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$, $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$.

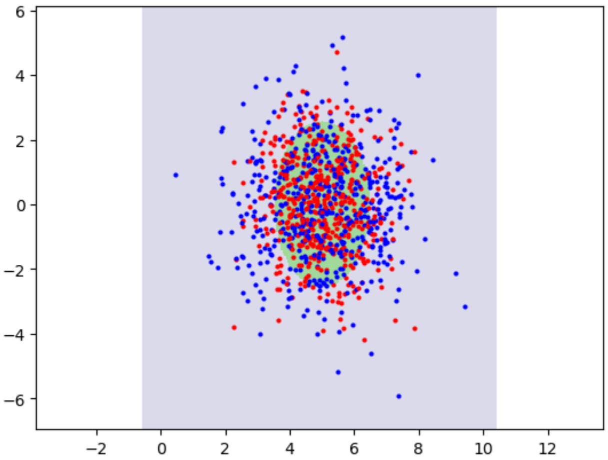


Figure 3: Problem 6(c). A elliptical boundary

7. *Experiments with least-squares regression.*

(a)

- The feature `MedInc` is the most highly correlated feature with`MedHouseVal`, since the correlation between them is 0.69.

- The pair of features `AveBedrms` and `AveRooms` are the most positively correlated, since the correlation between them is 0.85.

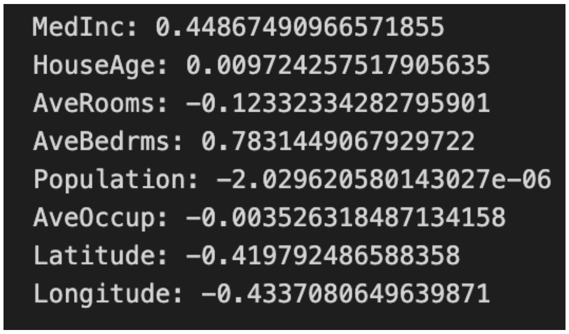
- The pair of features `Longitude` and `Latitude` are the most negatively correlated, since the correlation between them is -0.92.

(b)

- There are 16512 points in the training set, and 4128 points in the test set.

- The single best value to predict for the test set is 2.06, while the resulting MSE is 1.31.

(c) The coefficients of the linear model is:



```
MedInc: 0.44867490966571855
HouseAge: 0.009724257517905635
AveRooms: -0.12332334282795901
AveBedrms: 0.7831449067929722
Population: -2.029620580143027e-06
AveOccup: -0.003526318487134158
Latitude: -0.419792486588358
Longitude: -0.4337080649639871
```

Figure 4: Problem 7(c). Coefficients of the linear model

And the MSE on the test set is 0.5559.

(d) The MSE in this case is 0.9788.

(e) The variable `MedInc` is the best one to choose, and the resulting MSE is 0.7091.

8. *Discovering relevant features in regression.*

(a) My Strategy is to use Lasso to do feature selection, i.e. to use regularization with L1 penalty. Because Lasso can make the coefficient matrix sparse, thus selecting the most important features among a large amount of features.

(b) The ten features that I identify is: 2, 3, 5, 7, 11, 13, 17, 19, 23, 27.