Riddhi Gupta

23B3989

$(y_i - \hat{y}_i)^2$

Debraj Chakraborty

EE353 Mid-Sem

Date: 16-Sep-2025

Full Marks: 40 Time: 2 hrs

1. (10 points) Consider the simple linear regression model

$$y_i = 2 + \beta_1 x_i + \varepsilon_i, \qquad \varepsilon_i \overset{\text{i.i.d.}}{\sim} (0, \sigma^2), \quad i = 1, \ldots, n.$$

Use the following data generated from the above model:

$x_i (y_i - \bar{y})$

i	1	2	3	4
x_i	1	2	3	4
y_i	3	4	4	3

(a) (3 points) Find the least-squares estimator $\widehat{\beta}_1$ of the slope. Show all calculations.

(b) (3 points) Using an appropriate estimator for σ^2 computed from the data, calculate the standard error $\widehat{\text{SE}}(\widehat{\beta}_1)$.

(c) (3 points) Construct a 95% confidence interval for β_1. Use the appropriate t-value using the table below.

(d) (1 point) Compare, without explicit computation, the 95% CI that would have resulted if the intercept was also not known, with the CI computed above in part (c).

Table: Selected critical values of the t-distribution

df	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	12.71	31.82	63.66
2	4.303	6.965	9.925
3	3.182	4.541	5.841
4	2.776	3.747	4.604
5	2.571	3.365	4.032
6	2.447	3.143	3.707
∞	1.960	2.326	2.576

2. (10 points) Consider a binary response variable $Y \in \{0, 1\}$ and a single predictor $X \in \mathbb{R}$. Suppose we model

$$\Pr(Y = 1 \mid X = x) = \pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}.$$

(a) (1 point) Compute the log-likelihood $\ell(\beta_0, \beta_1)$ for n independent observations (x_i, y_i).

(b) (3 points) Compute the score equations (the first derivatives of the log-likelihood with respect to β_0 and β_1).

(c) (6 points) Compute the maximum likelihood estimates of β_0 and β_1 for the following dataset, using the score equations derived above:

i	1	2	3	4
x_i	-1	-1	1	1
y_i	1	0	1	0

3. (10 points) Consider a binary classification problem with classes $Y \in \{1, 2\}$ and three features $X = (X_1, X_2, X_3)$. The features are assumed to be conditionally independent given the class. The class priors are $\hat{\pi}_1 = 0.2$ and $\hat{\pi}_2 = 0.8$.

   Let $\hat{f}_{kj}(x)$ denote the estimated conditional density (for continuous features) or probability mass function (for categorical features) of feature X_j evaluated at x, given that the class is $Y = k$, i.e.

$$\hat{f}_{kj}(x) \approx \Pr(X_j = x \mid Y = k), \qquad k = 1, 2, \;\; j = 1, 2, 3.$$

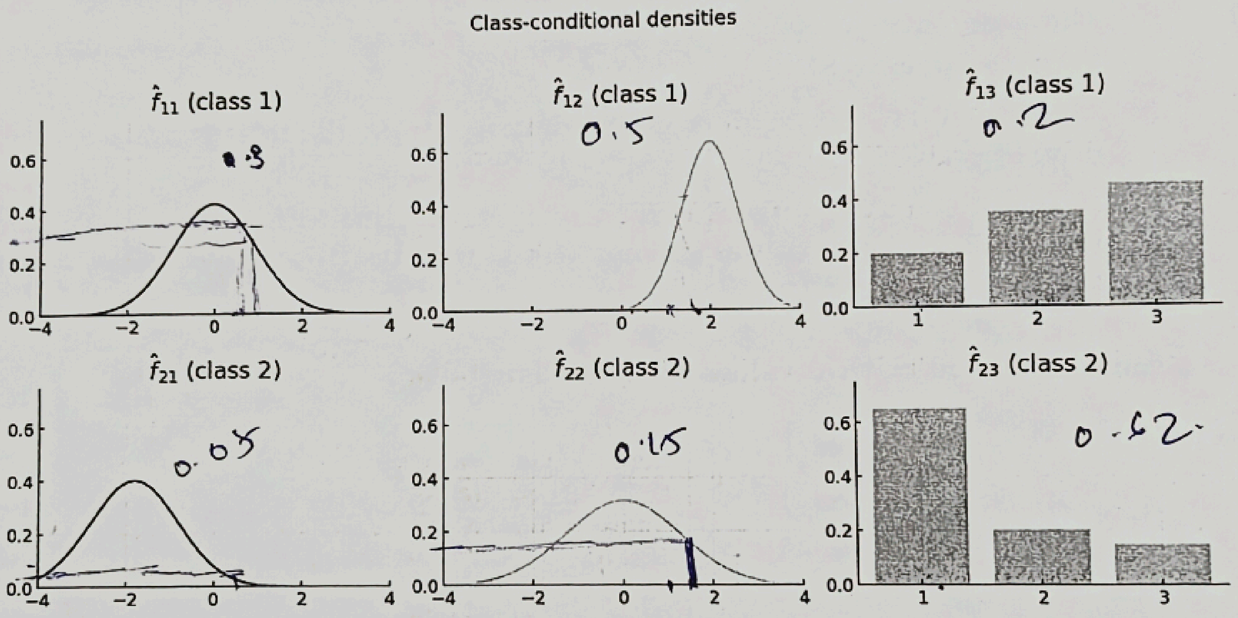

Figure 1: Estimated $\hat{f}_{kj}$'s

   (a) (8 points) For $x^* = (0.4, 1.5, 1)$, obtain the approximate posterior probabilities

$$\Pr(Y = 1 \mid X = x^*) \quad \text{and} \quad \Pr(Y = 2 \mid X = x^*).$$

   (b) (2 points) State the predicted class for x^*. Briefly explain your reasoning.

4. (10 points) Figure 2 is the output from a python program performing FDA on the two class data set plotted in the same figure. The program also contains the following lines of code:

========= CODE ============

0 mg

p :

3.889
1.76777

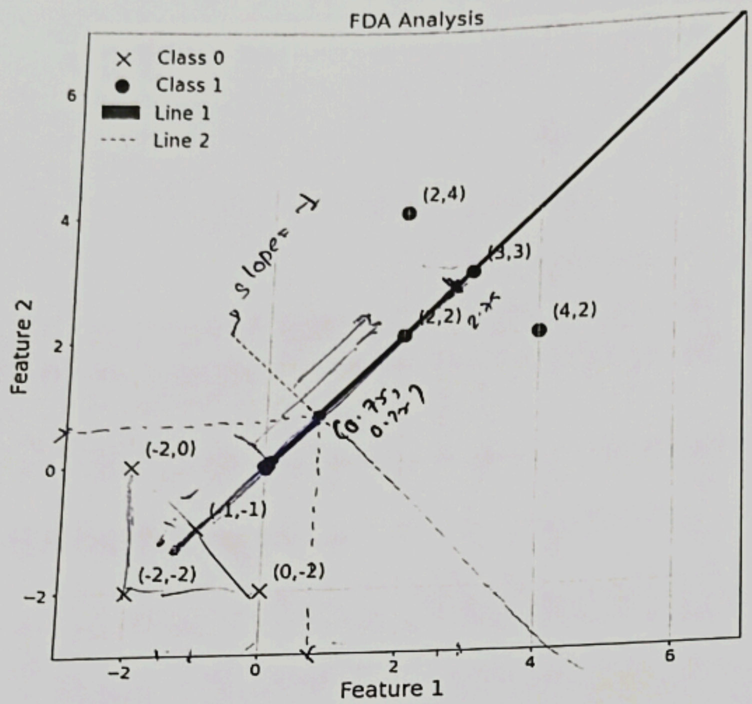

Figure 2: FDA Analysis

```
import numpy as np
n_samples_per_class = 4
X1 = np.array([[-2, -2],[-1, -1],[-2, 0],[0, -2]], dtype=float)
X2 = np.array([[2, 2],[3, 3],[2, 4],[4, 2]], dtype=float)
X = np.vstack([X1, X2])
mean0 = np.mean(X1, axis=0)
mean1 = np.mean(X2, axis=0)
global_mean = np.mean(X, axis=0)
diff0 = mean0 - global_mean
diff1 = mean1 - global_mean
# Important Matrices
Sb = n_samples_per_class * (np.outer(diff0, diff0) + np.outer(diff1, diff1))
S0 = np.zeros((2, 2))
S1 = np.zeros((2, 2))
for x in X1:
    diff = x - mean0
    S0 += np.outer(diff, diff)
for x in X2:
    diff = x - mean1
    S1 += np.outer(diff, diff)
Sw = S0 + S1
print(Sb)
print(Sw)
```

(−1.25, −1.25)
(2.75, 2.75)
(0.75, 0.75)
→ −2
→ 2

====================

This code generates the following output:

========= CODE OUTPUT ===========

```
[[32. 32.]
 [32. 32.]]
[[ 5.5 -2.5]
 [-2.5  5.5]]
```

====================

(a) (5 points) How is line 1 defined? Compute the equation of line 1 from the definition.

(b) (5 points) How is line 2 defined? Compute the equation of line 2 from the definition.

List of Formula

- Standard Simple Linear Reg.:

$$\hat{\beta}_1 = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2}, \; \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

$$SE(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_i (X_i - \bar{X})^2}}$$

$$SE(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_i (X_i - \bar{X})^2} \right)}$$

- LDA:

$$P(X|Y=k) \sim \mathcal{N}(\mu_k, \Sigma)$$

$$f_k(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1} (x-\mu_k)}$$

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k$$

- QDA:

$$P(X|Y=k) \sim \mathcal{N}(\mu_k, \Sigma_k)$$

$$f_k(x) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)}$$

$$\delta_k(x) = -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k) + \log \pi_k$$

- NB:

$$f_k(x) = \prod_{j=1}^{p} \frac{1}{\sqrt{2\pi\sigma_{kj}^2}} e^{-\frac{(x_j - \mu_{kj})^2}{2\sigma_{kj}^2}}$$

$$P(Y=k|X=x) \propto \pi_k f_k(x)$$

- FDA:

$$S_W = \sum_{k=1}^{K} \sum_{x_i \in C_k} (x_i - \mu_k)(x_i - \mu_k)^T$$

$$S_B = \sum_{k=1}^{K} (\mu_k - \mu)(\mu_k - \mu)^T$$

$$J(W) = \frac{|W^T S_B W|}{|W^T S_W W|}, \quad S_B v = \lambda S_W v$$