UNIVERSIDAD La Salle

Facultad de Negocios

Estadística Multivariada

Examen Final

1. LDA y QDA

Ejercicio 1. Supongamos que $X_1 \sim N_r(\mu_1, \Sigma_{X,X})$ y $X_2 \sim N_r(\mu_2, \Sigma_{X,X})$ son independientes. Considere el estadístico

$$\frac{\left(\mathbb{E}[a^t X_1] - \mathbb{E}[a^t X_2]\right)^2}{\mathrm{Var}\,(a^t X_1 - a^t X_2)}$$

como función de a. Demuestra que $a \propto \Sigma_{X,X}^{-1}(\mu_1 - \mu_2)$ maximiza el estadístico usando multiplicadores de Lagrange.

Ejercicio 2. Diga, explicando sus conclusiones, si las siguientes afirmaciones son falsas o verdaderas. Justifique adecuadamente su respuesta.

(1) Si el límite de decisión de Bayes es lineal, ¿esperamos que LDA o QDA funcionen mejor en el conjunto de entrenamiento?

(2) Si el límite de decisión de Bayes no es lineal, ¿esperamos que LDA o QDA funcionen mejor en el conjunto de entrenamiento?

(3) En general, conforme el tamaño de la muestra n aumenta, ¿esperamos que la precisión de la predicción de la prueba de QDA en relación con LDA mejore, disminuya o se mantenga sin cambios? ¿Por qué ?

(4) Si el límite de decisión de Bayes para un problema determinado es lineal, probablemente lograremos una tasa de error de prueba superior usando QDA en lugar de LDA porque QDA es lo suficientemente flexible como para modelar un límite de decisión lineal.

Ejercicio 3. Supongamos que se tienen dos variables aleatorias X_1 y X_2. Tomemos $X_3 = X_1^2, X_4 = X_2^2$ y $X_5 = X_1 X_2$. Dibuja las fronteras de LDA. Toma una base de datos de cualquier API y verifica que, efectivamente, se tienen las fronteras dibujadas. $\longrightarrow$ dibujar las curvas de nivel $\mathbb{R}^2 \longrightarrow \mathbb{R}$

Ejercicio 4. Considere algún conjunto de datos tomado de Kaggle [1]. **Nota: El conjunto de datos debe ser distinto para todos**. Define $X_5 = X_1/X_2$ y $X_6 = X_3/X_4$.

(1) ¿Qué mide X_5 y X_6?

(2) Toma la siguiente transformación: $\log(X_5)$ y $\log(X_6)$. Grafica los datos transformados y realiza un LDA.

(3) Realiza lo anterior para QDA.

(4) Generaliza para n cualquiera.

Ejercicio 5. Suponga que $X_1 \sim N_n(\mu_1,\ \Sigma_1)$ y que $X_2 \sim N_n(\mu_2,\ \Sigma_2)$. Suponga que $\Sigma_1 \neq \Sigma_2$. Demuestra que la log-verosimilitud del cociente de discriminación es de la forma

$$Q(x) = \beta_0 + \beta^{Tr} x + x^{Tr} O x$$

en donde tendrás qué encontrar a β_0, β y a O.

Ejercicio 6. Suponga que se coleccionan datos de un cierto grupo con variables aleatorias:

[1] https://www.kaggle.com

Universidad La Salle México

Tarea 1

Equipo:
Isabel Ruiz Mijangos
Alondra Vega Lázaro
Alejandro Zamora Reyes
Nadia Paola Jiménez Flores

Estadística multivariada

15/febrero/2024

UNIVERSIDAD La Salle

Facultad de Negocios

Estadística Multivariada

Examen Final

1. LDA y QDA

Ejercicio 1. Supongamos que $X_1 \sim N_r(\mu_1, \Sigma_{X,X})$ y $X_2 \sim N_r(\mu_2, \Sigma_{X,X})$ son independientes. Considere el estadístico

$$\frac{\left(\mathbb{E}[a^t X_1] - \mathbb{E}[a^t X_2]\right)^2}{\mathrm{Var}\left(a^t X_1 - a^t X_2\right)}$$

como función de a. Demuestra que $a \propto \Sigma_{X,X}^{-1}(\mu_1 - \mu_2)$ maximiza el estadístico usando multiplicadores de Lagrange.

Ejercicio 2. Diga, explicando sus conclusiones, si las siguientes afirmaciones son falsas o verdaderas. Justifique adecuadamente su respuesta.

(1) Si el límite de decisión de Bayes es lineal, ¿esperamos que LDA o QDA funcionen mejor en el conjunto de entrenamiento?

(2) Si el límite de decisión de Bayes no es lineal, ¿esperamos que LDA o QDA funcionen mejor en el conjunto de entrenamiento?

(3) En general, conforme el tamaño de la muestra n aumenta, ¿esperamos que la precisión de la predicción de la prueba de QDA en relación con LDA mejore, disminuya o se mantenga sin cambios? ¿Por qué ?

(4) Si el límite de decisión de Bayes para un problema determinado es lineal, probablemente lograremos una tasa de error de prueba superior usando QDA en lugar de LDA porque QDA es lo suficientemente flexible como para modelar un límite de decisión lineal.

Ejercicio 3. Supongamos que se tienen dos variables aleatorias X_1 y X_2. Tomemos $X_3 = X_1^2, X_4 = X_2^2$ y $X_5 = X_1 X_2$. Dibuja las fronteras de LDA. Toma una base de datos de cualquier API y verifica que, efectivamente, se tienen las fronteras dibujadas.

Ejercicio 4. Considere algún conjunto de datos tomado de Kaggle [1]. **Nota: El conjunto de datos debe ser distinto para todos**. Define $X_5 = X_1/X_2$ y $X_6 = X_3/X_4$.

(1) ¿Qué mide X_5 y X_6?

(2) Toma la siguiente transformación: $\log(X_5)$ y $\log(X_6)$. Grafica los datos transformados y realiza un LDA.

(3) Realiza lo anterior para QDA.

(4) Generaliza para n cualquiera.

Ejercicio 5. Suponga que $X_1 \sim N_n(\mu_1,\ \Sigma_1)$ y que $X_2 \sim N_n(\mu_2,\ \Sigma_2)$. Suponga que $\Sigma_1 \neq \Sigma_2$. Demuestra que la log-verosimilitud del cociente de discriminación es de la forma

$$Q(x) = \beta_0 + \beta^{Tr} x + x^{Tr} O x$$

en donde tendrás qué encontrar a β_0, β y a O.

Ejercicio 6. Suponga que se coleccionan datos de un cierto grupo con variables aleatorias:

[1] https://www.kaggle.com

(1) X_1 horas de estudio.

(2) X_2 promedio sobre 5.

(3) X_3 recibe una calificación de 10.

Suponga que $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.005$ y $\hat{\beta}_2 = 1$.

(1) ¿Qué significan tales estimaciones para $(\beta_0, \beta_1, \beta_2)$.

(2) Estima la probabilidad de que un estudiante que estudia 40 horas a la semana y tenga promedio de 3.5, obtenga una calificación de 10.

(3) ¿Cuántas horas debe estudiar para obtener un 50% de posibilidades de obtener un 10.

Ejercicio 7. Suponga que una observación en el k-ésimo grupo se toma a partir de una distribución $N(\mu_k, \sigma^2)$. Demuestra que el clasificador bayesiano asigna una observación a la clase para la cual la función discriminante se maximiza.

Ejercicio 8. Supongamos que deseamos predecir si una determinada acción emitirá un dividendo este año ("Sí" o "No") basándose en X, el porcentaje de beneficio del año pasado. Examinamos un gran número de empresas y descubrimos que el valor medio de X para las empresas que emitieron dividendos fue de 10, mientras que el valor medio para las que no lo hicieron fue 0. Además, la varianza de X para estas dos conjuntos de empresas fue 36. Finalmente, el 80% de las empresas emitieron dividendos. Suponiendo que X sigue una distribución normal, prediga la probabilidad de que una empresa emita un dividendo este año dado que su beneficio porcentual fue $X = 4$ el año pasado.

Ejercicio 9. Consideremos los datos de The Insurance Company Benchmark, que pueden descargarse de

$$kdd.ics.uci.edu/databases/tic$$

. Hay 86 variables sobre datos de uso de productos y datos sociodemográficos derivados de los códigos postales de los clientes de una compañía de seguros. Hay un conjunto de aprendizaje ticdata2000.txt de 5,822 clientes y un conjunto de prueba ticeval2000.txt de 4,000 clientes. Los clientes del conjunto de aprendizaje se clasifican en dos clases, dependiendo de si contrataron una póliza de seguro de caravana. El problema es predecir quién del conjunto de prueba estaría interesado en comprar una póliza de seguro de caravana. Utilice cualquiera de los métodos de clasificación en los datos de aprendizaje y, a continuación, aplíquelos a los datos de prueba. Compare sus predicciones para el conjunto de prueba con las del archivo tictgts2000.txt y calcule la tasa de error del conjunto de prueba. ¿Qué variables son más útiles para predecir la contratación de un seguro de caravana?

Ejercicio 1. Supongamos que $X_1 \sim N_r(\mu_1, \Sigma_{X,X})$ y $X_2 \sim N_r(\mu_2, \Sigma_{X,X})$ son independientes. Considere el estadístico

$$\frac{\left(\mathbb{E}[a^t X_1] - \mathbb{E}[a^t X_2]\right)^2}{\mathrm{Var}\left(a^t X_1 - a^t X_2\right)}$$

como función de a. Demuestra que $a \propto \Sigma_{X,X}^{-1}(\mu_1 - \mu_2)$ maximiza el estadístico usando multiplicadores de Lagrange.

$$\frac{(E[a^t X_1] - E[a^t X_2])^2}{Var(a^t X_1 - a^t X_2)}$$

$$E[a^t X_1] = a^t E[X_1] = a^t \mu_1$$

$$E[a^t X_2] = a^t E[X_2] = a^t \mu_2$$

$$Var(a^t X_1 - a^t X_2) = var(a^t X_1) + Var(a^t X_2) = 2\,Var(a^t X_1)$$

$$\text{estadístico} = \frac{(a^t \mu_1 - a^t \mu_2)^2}{a^t \Sigma_{x,x}\, a}$$

con lagrange

$$L(a, \lambda) = (a^t \mu_1 - a^t \mu_2)^2 - \lambda (a^t a - 1)$$

$$\frac{\partial L}{\partial a} = 2(a^t \mu_1 - a^t \mu_2)(\mu_1 - \mu_2) - 2\lambda a$$

$$2(a^t \mu_1 - a^t \mu_2)(\mu_1 - \mu_2) - 2\lambda a = 0$$

$$2(a^t \mu_1 - a^t \mu_2)(\mu_1 - \mu_2) = 2\lambda a$$

$$(a^t \mu_1 - a^t \mu_2)(\mu_1 - \mu_2) = \lambda a$$

$$a^t (\mu_1 - \mu_2)(\mu_1 - \mu_2) = \lambda a$$

$$a^t (\mu_1 - \mu_2)^2 = \lambda a$$

Ejercicio 2. Diga, explicando sus conclusiones, si las siguientes afirmaciones son falsas o verdaderas. Justifique adecuadamente su respuesta.

(1) Si el límite de decisión de Bayes es lineal, ¿esperamos que LDA o QDA funcionen mejor en el conjunto de entrenamiento?

(2) Si el límite de decisión de Bayes no es lineal, ¿esperamos que LDA o QDA funcionen mejor en el conjunto de entrenamiento?

(3) En general, conforme el tamaño de la muestra n aumenta, ¿esperamos que la precisión de la predicción de la prueba de QDA en relación con LDA mejore, disminuya o se mantenga sin cambios? ¿Por qué ?

(4) Si el límite de decisión de Bayes para un problema determinado es lineal, probablemente lograremos una tasa de error de prueba superior usando QDA en lugar de LDA porque QDA es lo suficientemente flexible como para modelar un límite de decisión lineal.

(1) Esperamos que LDA funcione mejor, ya que tiene una mejor estructura lineal para capturar los datos, en QDA puede haber un sobreajuste en los datos, ya que puede ser demasiado flexible y ajustarse demasiado a los datos.

(2) En este caso QDA funcionaría mejor ya que es más flexible porque no supone homogeneidad, por lo que puede capturar relaciones no lineales entre las características y las clases.

(3) Mejore, ya que QDA es mucho más flexible y puede ajustar de una mejor manera a una muestra de tamaño n, incluso si esta aumentara, por la naturalidad del QDA, ya que son funciones cuadráticas.

(4) Verdadero, ya que el QDA sin problemas se ajusta a una frontera lineal.

Ejercicio 5. Suponga que $X_1 \sim N_n(\mu_1,\ \Sigma_1)$ y que $X_2 \sim N_n(\mu_2,\ \Sigma_2)$. Suponga que $\Sigma_1 \neq \Sigma_2$. Demuestra que la log-verosimilitud del cociente de discriminación es de la forma

$$Q(x) = \beta_0 + \beta^{Tr}x + x^{Tr}Ox$$

en donde tendrás qué encontrar a β_0, β y a O.

QDA $\to \Sigma_1 \neq \Sigma_2 \to \log\left(\frac{f_1 \pi_1}{f_2 \pi_2}\right) =$ cuadratica

$$f_i(x) = \frac{}{\sqrt{(2\pi)^n |\det \Sigma_i|}} \exp\left\{ -\frac{1}{2} (x-\mu_i) \Sigma_i^{(-1)} (x-\mu_i)^{tr} \right\}$$

$$f_1 = \frac{1}{\sqrt{(2\pi)^n |\det \Sigma_1|}} \exp\left\{ -\frac{1}{2} (x-\mu_1) \Sigma_1^{-1} (x-\mu_1)^t \right\}$$

$$f_2 = \frac{1}{\sqrt{(2\pi)^n |\det \Sigma_2|}} \exp\left\{ -\frac{1}{2} (x-\mu_2) \Sigma^{-1} (x-\mu_2)^t \right\}$$

$$\frac{f_1(x)\,\pi_1}{f_2(x)\,\pi_2} = \frac{\frac{\pi_1}{\sqrt{(2\pi)^n |\det \Sigma_1|}} \exp\left\{ -\frac{1}{2} (x-\mu_1) \Sigma_1^{-1} (x-\mu_1)^t \right\}}{\frac{\pi_2}{\sqrt{(2\pi)^n |\det \Sigma_2|}} \exp\left\{ -\frac{1}{2} (x-\mu_2) \Sigma^{-1} (x-\mu_2)^t \right\}}$$

$$= \frac{\pi_1}{\pi_2} \cdot \frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}} \cdot \exp\left(-\frac{1}{2} (x-\mu_1) \Sigma_1^{-1} (x-\mu_1)^t + \frac{1}{2} (x-\mu_2) \Sigma^{-1} (x-\mu_2)^t \right)$$

$$= \log\left(\frac{f_1(x)\,\pi_1}{f_2(x)\,\pi_2}\right) = \log\left(\frac{\pi_1}{\pi_2}\right) + \log\left(\frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}}\right) - \frac{1}{2} (x-\mu_1) \Sigma_1^{-1} (x-\mu_1)^t$$
$$+ \frac{1}{2} (x-\mu_2) \Sigma^{-1} (x-\mu_2)^t$$

$$= \log(\pi_1) - \log(\pi_2) + \frac{1}{2}\log(|\Sigma_2|) - \frac{1}{2}\log(|\Sigma_1|) + (\mu_2 - \mu_1)^t x \quad \Sigma_1^{-1}$$
$$-\frac{1}{2} (\mu_2 - \mu_1)^t \Sigma_2^{-1} (\mu_2 - \mu_1)$$

si hacemos $\beta = \Sigma_2^{-1}(\mu_2 - \mu_1)$ entonces

$$= \log(\pi_1) - \log(\pi_2) + \frac{1}{2}\log(|\Sigma_2|) - \frac{1}{2}\log(|\Sigma_1|) + \beta^t X$$
$$-\frac{1}{2}(\mu_2 - \mu_1)^t \Sigma_2^{-1}(\mu_2 - \mu_1)$$

si hacemos $\beta_0 = \log\left(\frac{\pi_1}{\pi_2}\right) + \frac{1}{2}\log\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right)$

$$= \beta_0 + \beta^t X - \frac{1}{2}(\mu_2 - \mu_1)^t \Sigma_2^{-1}(\mu_2 - \mu_1)$$

Para O necesitamos que sea simetrica y positiva definida y así $O = -\frac{1}{2}(\Sigma_2^{-1} - \Sigma_1^{-1})$

$$= \beta_0 + \beta^t X + x^t O x$$

así

$$Q(x) = \beta_0 + \beta^t X + x^t O x$$

Ejercicio 6. Suponga que se coleccionan datos de un cierto grupo con variables aleatorias:

(1) X_1 horas de estudio.
(2) X_2 promedio sobre 5.
(3) X_3 recibe una calificación de 10.

Suponga que $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.005$ y $\hat{\beta}_2 = 1$.

(1) ¿Qué significan tales estimaciones para $(\beta_0, \beta_1, \beta_2)$.
(2) Estima la probabilidad de que un estudiante que estudia 40 horas a la semana y tenga promedio de 3.5, obtenga una calificación de 10.
(3) ¿Cuántas horas debe estudiar para obtener un 50% de posibilidades de obtener un 10.

$$L(x) = \beta_0 + \beta^T x$$

(2)

$$\hat{L}(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

$$p(\pi_1 | x) = \frac{e^{\hat{L}(x)}}{1 + e^{\hat{L}(x)}} = \frac{e^{-6 + .005(40) + 3.5}}{1 + e^{-6 + .005(40) + 3.5}} = 0.0911 \approx 9.11\%$$

(3)

$$0.50 = \frac{1}{1 + e^{+6 - .005(x_1) - 3.5}}$$

$$0.50(1 + e^{+6 - .005(x_1) - 3.5}) = 1$$

$$\log(e^{+6 - .005(x_1) - 3.5}) = \log(1/.5 - 1) \implies 500 \text{ horas}$$

$$-.005(x_1) = -6 + 3.5$$

$$x_1 = 500$$

(1) X_1 horas de estudio.

(2) X_2 promedio sobre 5.

(3) X_3 recibe una calificación de 10.

Suponga que $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.005$ y $\hat{\beta}_2 = 1$.

(1) ¿Qué significan tales estimaciones para $(\beta_0, \beta_1, \beta_2)$.

(2) Estima la probabilidad de que un estudiante que estudia 40 horas a la semana y tenga promedio de 3.5, obtenga una calificación de 10.

(3) ¿Cuántas horas debe estudiar para obtener un 50% de posibilidades de obtener un 10.

β_1: indica como es la variable x (horas de estudio) afecta al logaritmo de la odds de tener una calificación de 10. En este caso, = 0.005 sugiere que por cada unidad adicional de horas de estudio, el logaritmo de la odds. aumenta en 0.005.

β_2: Similar a β_1, pero la variable x_2 (promedio sobre cinco). Un valor de $\beta_2 = 1$ indica que un aumento de unidad en el promedio sobre cinco, se asocia con un aumento de uno en el logaritmo de obtener una calificación de 10.

En conclusión, dichas estimaciones proporcionan información sobre cada variable predictora contribuye a logaritmo de obtener una calificación de 10 en el modelo en regresión.

```
import numpy as np

# Coeficientes estimados
beta_0 = -6
beta_1 = 0.005
beta_2 = 1

def logistic_function(x1, x2):
    # Función logística
    log_odds = beta_0 + beta_1 * x1 + beta_2 * x2
    return 1 / (1 + np.exp(-log_odds))

def estimate_probability(x1, x2):
    # Estimar la probabilidad usando la función logística
    return logistic_function(x1, x2)

def find_hours_for_50_percent_probability():
    # Resolver la ecuación logística para encontrar las horas de estudio necesarias para un 50%
    result = np.log(0.5 / (1 - 0.5)) - beta_0
    hours_needed = result / beta_1
    return hours_needed

# (2) Estimación de la probabilidad para un estudiante que estudia 40 horas y tiene un promedio de 3.5
probability = estimate_probability(40, 3.5)
print(f"Probabilidad de obtener una calificación de 10: {probability:.2f}")

# (3) Encontrar las horas necesarias para un 50% de posibilidades
hours_needed = find_hours_for_50_percent_probability()
print(f"Horas de estudio necesarias para un 50% de posibilidades: {hours_needed:.2f}")

```

Respuesta:

```
Probabilidad de obtener una calificación de 10: 0.09
Horas de estudio necesarias para un 50% de posibilidades: 1200.00
```

Ejercicio 8. Supongamos que deseamos predecir si una determinada acción emitirá un dividendo este año ("Sí" o "No") basándose en X, el porcentaje de beneficio del año pasado. Examinamos un gran número de empresas y descubrimos que el valor medio de X para las empresas que emitieron dividendos fue de 10, mientras que el valor medio para las que no lo hicieron fue 0. Además, la varianza de X para estas dos conjuntos de empresas fue 36. Finalmente, el 80% de las empresas emitieron dividendos. Suponiendo que X sigue una distribución normal, prediga la probabilidad de que una empresa emita un dividendo este año dado que su beneficio porcentual fue $X = 4$ el año pasado.

Y: v.a. si empresa emite dividendos o no $\begin{cases} 1 & \text{si emite} \\ 0 & \text{no emite} \end{cases}$

X: % de beneficio del año pasado

$X \sim$ Normal

$$P(Y|X) = \frac{P(X|Y)\cdot P(Y)}{P(X)}$$

Los que emiten dividendos

$\mu_Y = 10$

$\sigma^2_Y = 36$

Los que no emiten dividendos

$\mu_Z = 0$

$\mu_Z = 36$

Ahora

$$P(Y|X=4) = \frac{P(X=4|Y)\,P(Y)}{P(X=4)}$$

$$P(X=4|Y) = \frac{1}{\sqrt{2\pi(36)}}\, e^{-\frac{(4-10)^2}{2(36)}} = \frac{1}{6\sqrt{\pi}}\, e^{-1/2} = 0.1065$$

$$P(X=4|Z) = \frac{1}{\sqrt{2\pi(36)}}\, e^{-\frac{(4-0)^2}{2(36)}} = .01295$$

$$P(X=4) = P(X=4|Y)\cdot P(Y) + P(X=4|Z)\cdot P(Z)$$

$$= 0.1065(.80) + .01295(.2)$$

$$= .1114$$

Ahora

$$P(Y|X=4) = \frac{P(X=4|Y)\cdot P(Y)}{P(X=4)} = \frac{.1065(.8)}{.1114} = .7644$$

Ejercicio 7. Suponga que una observación en el k-ésimo grupo se toma a partir de una distribución $N(\mu_k, \sigma^2)$. Demuestra que el clasificador bayesiano asigna una observación a la clase para la cual la función discriminante se maximiza.

El clasificador asigna una observación a la clase con la mayor probabilidad, esto es:

$$p(\pi_K \mid x) = \frac{f_K(x)\,\pi_K}{\sum_i f_i(x)\,\pi_i} = \frac{\pi_K\, e\left\{-\frac{1}{2\sigma^2}(x-\mu_K)^2\right\}}{\sum_{i=1}^{K} \pi_i\, e\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}}$$

Con logaritmo natural tenemos:

$$\ln p_K(x) = \ln(\pi_K) - \frac{1}{2\sigma^2}(x-\mu_K)^2 - \ln \sum_{i=1}^{K} \pi_i\, e\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}$$

Es equivalente a encontrar K;

$$\therefore \quad \frac{\mu_K}{\sigma^2}x - \frac{\mu_K^2}{2\sigma^2} + \log \pi_K$$