



**UNIVERSIDAD NACIONAL DE INGENIERÍA**  
**FACULTAD DE INGENIERÍA MECÁNICA**  
Escuela Profesional de Ingeniería Mecatrónica

**UNI Machine Learning (MT-616-A)**

Paul Cárdenas Lizana

**Práctica No 2**

**(Sección A)**

1. **(4 puntos)** El objetivo del análisis discriminante lineal (LDA, e.g., discriminante de Fisher) es proporcionar una separación óptima de clases (con respecto a la media) y al mismo tiempo mantener muy pequeña la varianza de la muestra dentro la clase. Esto se formaliza maximizando la función,  $J(\theta)$ , con respecto a  $\theta$ , donde se define  $\mathbf{m}_1$  y  $\mathbf{m}_2$  como la media de los dos conjuntos de datos etiquetados,  $C_1$  y  $C_2$ ,

$$J(\theta) = \frac{\theta^T S_B \theta}{\theta^T S_W \theta}$$

donde  $S_B$  es la matriz de covarianza entre clases y  $S_W$  es la matriz de covarianza total dentro de la clase.

$$S_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T$$

$$S_W = \sum_{i=1,2} \sum_{n=1}^{l_i} (\mathbf{x}_n^i - \mathbf{m}_i)(\mathbf{x}_n^i - \mathbf{m}_i)^T$$

$l_i$  representa el número de ejemplos de la clase  $C_i$ . LDA intuitivamente trata de encontrar una proyección donde se maximiza la separación de las clases. Se pide

- a. Mostrar que maximizar  $J(\theta)$  es igual a reformular el problema como

$$\max \theta^T S_B \theta$$

Sujeto a la condición  $\theta^T S_W \theta = 1$

- b. Hallar el  $\theta$  que maximiza  $J(\theta)$

2. **(4 puntos)** Muestre que:

- a. Para un conjunto de datos linealmente separables, la solución de probabilidad máxima para el modelo de regresión logística se obtiene al encontrar un vector  $\theta$  cuyo límite de decisión  $\theta^T \phi(\mathbf{x}) = 0$  separa las clases y luego toma la magnitud de  $\theta$  hasta el infinito, donde  $\phi$  es una función base. (Hint: trata de usar casos extremos en python para graficar la función logística)
- b. La función logística satisface la propiedad  $\sigma(-\mathbf{a}) = 1 - \sigma(\mathbf{a})$  y que su inversa está dada por  $\sigma^{-1}(y) = \ln\{y/(1 - y)\}$ .



**UNIVERSIDAD NACIONAL DE INGENIERÍA**  
**FACULTAD DE INGENIERÍA MECÁNICA**  
Escuela Profesional de Ingeniería Mecatrónica

3. **(4 puntos)** Un hiperplano de separación óptimo separa dos clases y maximiza la distancia al punto más cercano de cualquiera de las clases. Esto no solo proporciona una solución única y óptima para el problema de separación, sino que también maximiza el margen entre las dos clases en los datos de entrenamiento. Este método propone un mejor rendimiento de clasificación. En este problema debes de maximizar

$$\max_{\theta, \theta_o, ||\theta||=1} R$$

Sujeto a la condición de

$$y_i(x_i^T \theta + \theta_o) \geq R \quad \forall \quad i = 1, \dots, N.$$

para obtener el hiperplano óptimo. El conjunto de condiciones garantiza que todos los puntos estén al menos a una distancia positiva  $R$  del límite de decisión definido por  $\theta$  y  $\theta_o$ , y así buscar el  $R$  más grande con sus parámetros asociados. Este es el problema primal. Como se demostró en clase usar los 3 intentos para transformarlo a un problema convexo (Problema dual)

4. **(8 puntos) Regresión logística para sistema médico de recomendación.** Un ataque al corazón es una emergencia médica muy delicada y usualmente ocurre cuando un coágulo de sangre bloquea el flujo de sangre al corazón. Sin sangre, el tejido pierde oxígeno y muere. Los síntomas incluyen opresión o dolor en el pecho, cuello, espalda o brazos, así como fatiga, aturdimiento, latidos cardíacos anormales y ansiedad. El tratamiento abarca desde cambios en el estilo de vida y rehabilitación cardíaca hasta medicamentos, stents y cirugía de bypass. En este problema vas a ayudar a un médico a dar una respuesta informada e inteligente al predecir si el paciente está cerca de tener un ataque al corazón (porcentaje). Los datos tiene 200 datos de entrenamiento de pacientes con 13 atributos medidos experimentalmente y cada uno etiquetado debidamente con presencia o ausencia de enfermedad. Tu tarea es utilizar todo lo que se ha aprendido en el curso de ML para desarrollar un sistema médico de recomendación inteligente.

Tu nota va depender de la precisión de tu algoritmo de ML en 50 datos no vistos (no dados)

[https://github.com/PCL-AI/MT616\\_2018\\_2/blob/master/Lab03/Lab3.ipynb](https://github.com/PCL-AI/MT616_2018_2/blob/master/Lab03/Lab3.ipynb)

Atributos:

1. Edad: edad en años
2. Sexo:
  - Valor 1 = masculino
  - Valor 0 = femenino
3. cp: tipo de dolor de pecho (La angina típica es la incomodidad que se observa cuando el corazón no recibe suficiente sangre u oxígeno.)



**UNIVERSIDAD NACIONAL DE INGENIERÍA**  
**FACULTAD DE INGENIERÍA MECÁNICA**  
Escuela Profesional de Ingeniería Mecatrónica

- Valor 1: angina típica
  - Valor 2: angina atípica
  - Valor 3: dolor no anginal
  - Valor 4: asintomático
4. Trestbps: presión arterial en reposo (en mm Hg al ingreso al hospital)
5. col: colesterol en la sangre mg/dl
6. fbs: (azúcar en la sangre en ayunas > 120 mg / dl)
- Valor 1 = verdadero
  - Valor 0 = falso
7. restecg: (resultados reposo electrocardiográfico)
- Valor 0: normal
  - Valor 1: anomalía de la onda ST-T
  - Valor 2: muestra la hipertrofia ventricular izquierda probable o definitiva
8. thalach: máximo ritmo cardiaco alcanzado
9. exang: angina inducida por el ejercicio
- Valor 1 = si
  - Valor 0 = no
10. oldpeak: depresión del ST inducida por el ejercicio en relación con el descanso.
11. slope: la pendiente del segmento pico del ejercicio
- Valor 1: ascendente
  - Valor 2: plano
  - Valor 3: descendente
12. ca: número de vasos importantes (0-3) coloreados por fluoroscopia
13. thal: Tomografía de thallium
- Valor 3 = normal (sin puntos)
  - Valor 6 = defecto fijo (puntos durante el descanso y el ejercicio)
  - Valor 7 = defecto reversible (los puntos solo aparecen durante el ejercicio)
14. pred\_attribute: (predicción) diagnóstico de enfermedad cardíaca
- absent
  - present