

Auxiliar 9: Repaso C2 y PCA - Solución

P1. Clasificación generativa vs. discriminativa

a) Clasificación de postulaciones de trabajo y detección de outliers.

Recomendación: Usar un modelo generativo, como la estimación de densidad.

b) Clasificación de aplicaciones de software con poca data.

Recomendación: Usar un modelo generativo debido a la escasez de datos.

c) Clasificación de compañías con mucha data de entrenamiento.

Recomendación: Usar un modelo discriminativo ya que hay suficiente data.

P2. Clasificación de billetes con LDA

a) Estimación de la matriz de covarianza común.

Fórmula: $\Sigma = (n_V * \Sigma_V + n_F * \Sigma_F) / (n_V + n_F)$.

b) Suposiciones para LDA: distribuciones gaussianas y covarianza igual.

Regla de clasificación: $\delta_k(x) = x^T \Sigma^{-1} \mu_k - 1/2 \mu_k^T \Sigma^{-1} \mu_k + \log(\pi_k)$.

c) Clasificación del billete con largo 214.0 y diagonal 140.4 usando la regla de LDA.

P3. Kernel para Support Vector Machine (SVM)

a) Verificación de que el kernel es un Mercer kernel mostrando que es semidefinido positivo.

b) Mapeo de los datos al nuevo espacio de características.

c) Dibujar el plano de decisión para el SVM de margen máximo en el nuevo espacio de características.

d) Dibujar el plano de decisión en el espacio original.

P4. Kernel PCA

a) Mostrar que los vectores propios de Kernel PCA son combinaciones lineales de las características $\phi(x_i)$.

b) Implementación de Kernel PCA para datasets de juguete.