Tarea 05 - Reconocimiento de patrones

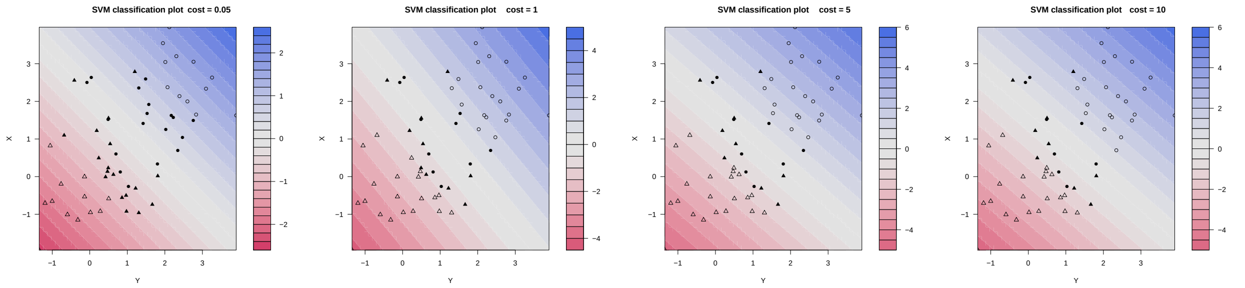
Giovanni Gamaliel López Padilla

Problema 01

Explora en https://colab.research.google.com/drive/1pwCqLdvxeqChzDG3MoFIgR7m6lsInKs5?usp=sharing el efecto de cambiar los parámetros en una SVM y convéncete que es de acuerdo a (congruente con) el funcional de costo de una SVM.

Experimento 1

¿Cuál es el efecto de cambiar el parámetro λ/γ (cost)?

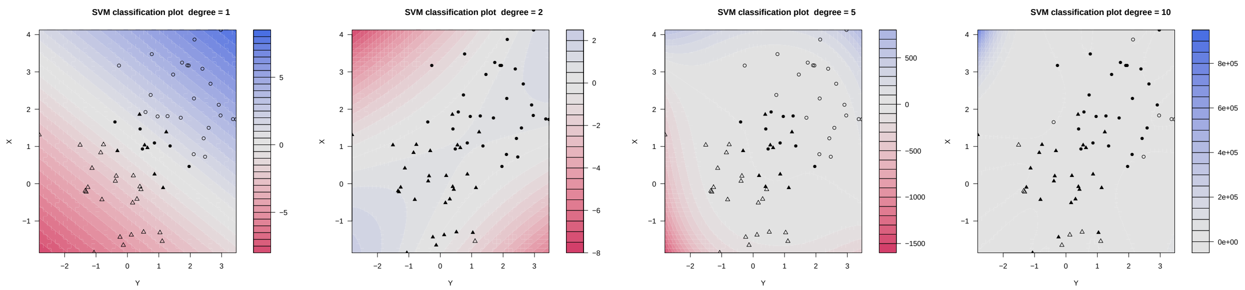


(a) (b) (c) (d)

Figura 1

Experimento 2

¿Cuál es el efecto de aumentar el grado de polinomio?

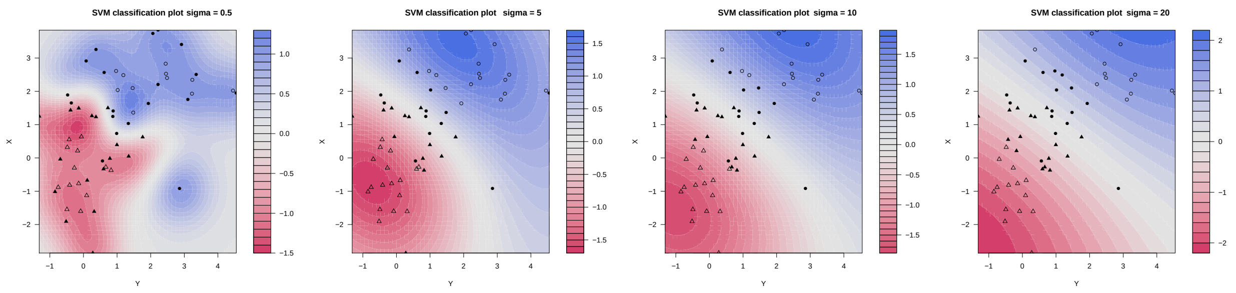


(a) (b) (c) (d)

Figura 2

Experimento 3

¿Cuál es el efecto de cambiar el parámetro de kernel de base radial?

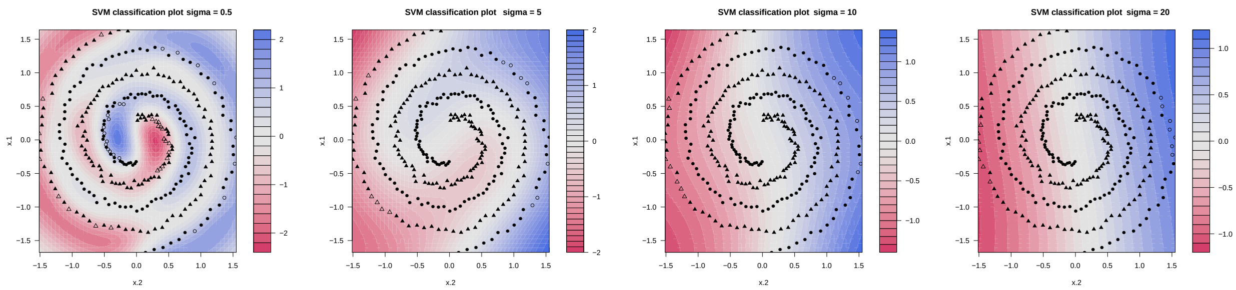


(a) (b) (c) (d)

Figura 3

Experimento 4

¿Cuál es el efecto de cambiar el parámetro σ del kernel de base radial?



(a) (b) (c) (d)

Figura 4

Problema 2

Vimos que minimizar $E(1 - Y(g(X))_+$ sobre g conduce al clasificador óptimo $\hat{y}(x) = sgn(g(x))$. Usando el mismo camino, muestra que se obtiene el mismo resultado para $E(exp(-Yg(X)))$

El problema se encuentra planteado en la ecuación 1.

$$\min_{g(x)} E(e^{-yg(x)}) \tag{1}$$

Por el teorema de números grandes, el problema de la ecuación 1 se puede resolver encontrando el valor de $g(x)$ para cada x dada. Esto puede ser visto en la ecuación 2.

$$\min_{g(x)} E_{Y|X=x}(e^{-yg(x)}) \tag{2}$$

Realizando el calculo explicito de la ecuación 2 se obtiene lo siguiente:

$$E_{Y|X=x}(exp(-yg(x))) = e^{-g(x)}P(Y = 1|X = x) + e^{g(x)}P(Y = -1|X = x)$$

por lo tanto:

$$
\begin{aligned}
\underset{g(x)}{min} E_{Y|X=x} &= \frac{\partial}{\partial g(x)} \left(e^{-g(x)} P(Y=1|X=x) + e^{g(x)} P(Y=-1|X=x) \right) \\
&= -e^{-g(x)} P(Y=1|X=x) + e^{g(x)} P(Y=-1|X=x)
\end{aligned}
$$

Encontrando el valor crítico del resultado anterior se obtiene que $g(x)$ es:

$$
\begin{aligned}
&- e^{-g(x)} P(Y=1|X=x) + e^{g(x)} P(Y=-1|X=x) = 0 \\
&- P(Y=1|X=x) + e^{2g(x)} P(Y=-1|X=x) = 0 \\
&e^{2g(x)} P(Y=-1|X=x) = P(Y=1|X=x) \\
&e^{2g(x)} = \frac{P(Y=1|X=x)}{P(Y=-1|X=x)} \\
&2g(x) = \log\left(\frac{P(Y=1|X=x)}{P(Y=-1|X=x)}\right) \\
&g(x) = \frac{1}{2}\left(\log(P(Y=1|X=x)) - \log(P(Y=-1|X=x))\right)
\end{aligned}
$$

por lo tanto, $g(x)$ tiene esta descrito por la ecuación 3.

$$
g(x) = \frac{1}{2}\left(\log(P(Y=1|X=x)) - \log(P(Y=-1|X=x))\right) \tag{3}
$$

Tomando el caso cuando $P(Y=1|X=x) > P(Y=-1|X=x)$, se tiene que $g(x) > 0$, por lo tanto $\hat{y}(x) = 1$. Por otro lado cuando $P(Y=-1|X=x) > P(Y=1|X=x)$, entonces $g(x) < 0$, por lo tanto $\hat{y}(x) = -1$. El comportamiento que presenta $g(x)$ es el mismo que tiene el clasificador óptimo bayesiano. Por lo que el problema planteado en la ecuación 1 condude al clasificador óptimo bayesiano.

Problema 3

Supongamos que (X,Y) cumplen los supuestos del clasificador binario LDA. Sin embargo, a partir de una muestra de (X,Y), alguien decida usar QDA (el clasificador bayesiano óptimo para el caso donde $X|Y=Y \sim \mathcal{N}(\mu_y, \Sigma_y)$, o sea no aprovechar que las covarianzas son iguales) y no LDA.

Experimento 1

¿Cómo se comparan el error de entrenamiento de QDA con el de LDA para este caso? No hay que hacer cálculos formales sino dar argumentos intuituvos

Problema 4

Este ejercicio es sobre el uso de métodos de clasificación para detectar billetes falsos:

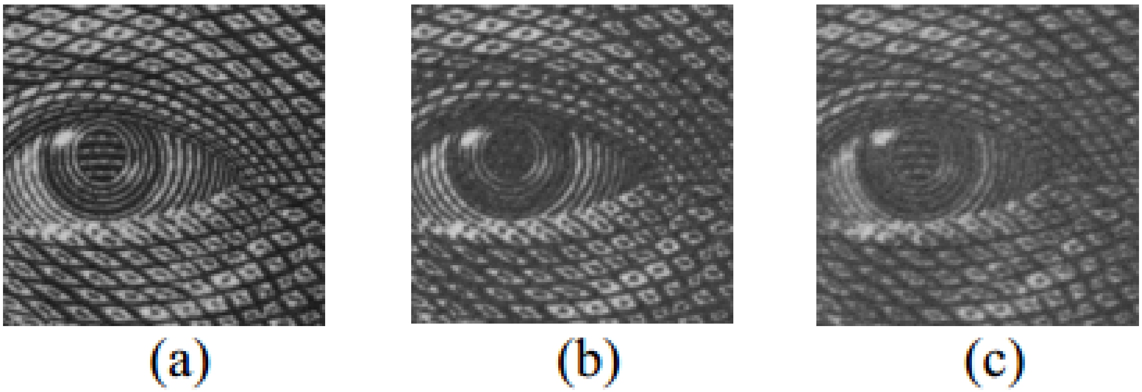


Figura 5: (a) (parde de) un billete de verdad, (b) billete falso de alta calidad, (c) billete falso de baja calidad.

En el paper que se anexa a la tarea se resume cada billete con cuatro caracteristicas (varianza, skewness, curtosis y entropía) extraidas de la forma del histograma de los coeficientes de la transformación de Wavelet. Los histogramas a continuación muestran como cambia la forma cuando el billete ya no es auténtico.

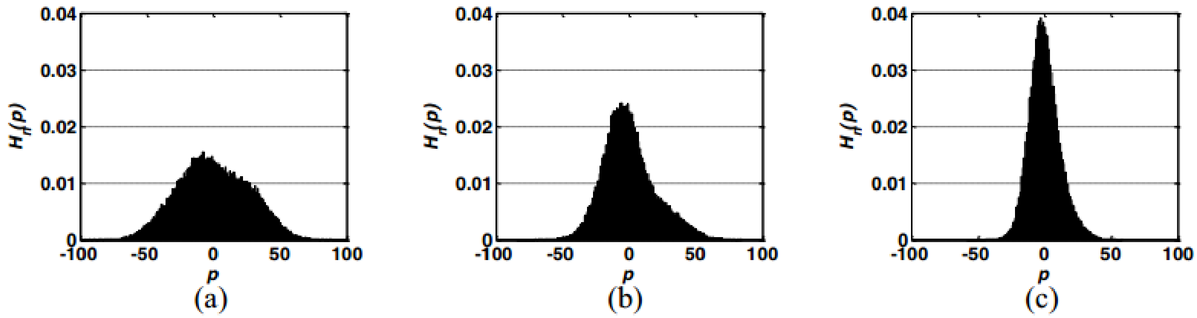


Figura 6: (a) histograma de los coeficientes de un billete de verdad, (b) billete falso de alta calidad, (c) billete falso de baja calidad.

Se anexo el conjunto de datos. La última columna indica si el billete es falso o no.

- Resume, visualiza y analiza los datos
- Construye algunos clasificadores interesantes basados en SVM (explora diferentes kerneles). Estima su poder predictivo, para eso divide muchas veces los datos en conjunto de prueba y de entrenamiento y cuenta falsos positivos y falsos negativos. Las instrucciones básicas de SVM para R y Python están al fianl de recpat4b.pdf