

## Tarea 5 Reconocimiento de Patrones

Entregar a más tardar el miércoles 27 de abril antes de las 10PM.

Se debe integrar todo en un solo pdf (sin comprimir).

### A. Cultura general (no hay que entregar nada)

1. Les recomiendo ampliamente la siguiente serie de videos por los autores del libro *Statistical Learning*  
<https://www.dataschool.io/15-hours-of-expert-machine-learning-videos/>

2. Se mencionó brevemente el **Representer Theorem** (recpat4b.pdf)

Para los interesados en la demostración, ver <https://www.youtube.com/watch?v=TMEgQzGJauU>.

Como intro (sin entrar en todos los detalles), lee la página 5 de recpat4b.pdf:

Toma por simplicidad el caso de un kernel lineal  $K(u, v) = \langle u, v \rangle = u^t v$ ,  $u, v \in X$ .

Sea  $\mathcal{H}$  el espacio de funciones lineales sobre  $X$ :  $g(x) = \beta_g^t x = K(\beta_g, x)$  y  $g(\cdot) = K(\beta_g, \cdot)$ . Definimos  $\langle g, h \rangle_{\mathcal{H}} = \beta_g^t \beta_h = K(\beta_g, \beta_h)$ .

Verifica que  $\|g\|_{\mathcal{H}} = \|\beta_g\|$  y  $\langle g, K(x_i, \cdot) \rangle_{\mathcal{H}} = g(x_i)$ .

Convéncete que la función de costo de una SVM cumple la forma del **Representer Theorem**.

### B. Preguntas cortas

1. 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 (recuerda: para ejecutar una celda: **Alt + Enter**).

(las preguntas están en el notebook)

2. Vimos que minimizar  $E(1 - Yg(X))_+$  sobre  $g$  conduce al clasificador Bayesiano óptimo  $\hat{y}(x) = \text{sgn}(g(x))$  (ver video en <https://youtu.be/bpZC-2zk0zk>)

Usando el mismo camino, muestra que se obtiene el mismo resultado para  $E \exp(-Yg(X))$ .

- 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.

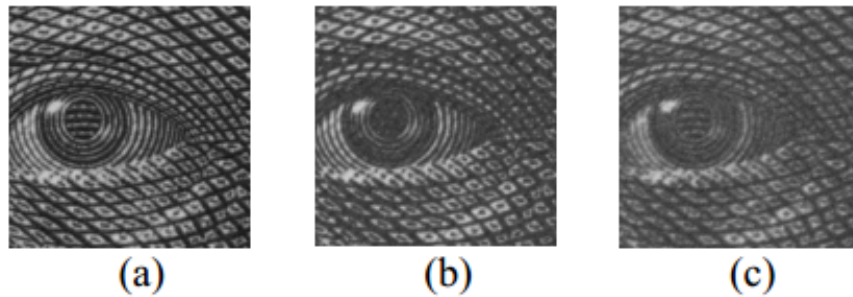
¿Cómo se comparan el error de entrenamiento de QDA con el de LDA para este caso?

Si uno tiene muchísimos datos, ¿cómo se comparan el error de prueba de QDA con el de LDA para este caso?

No hay que hacer cálculos formales sino dar argumentos intuitivos.

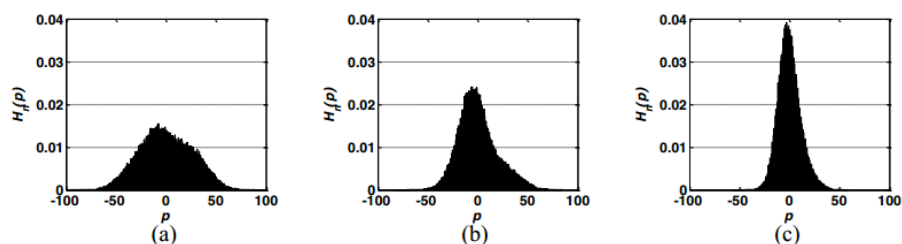
## C. Análisis de datos

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



(a) (parte 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 4 características (varianza, skewness, curtosis y entropía) extraídas 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.



(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 (sin hacer distinción entre falso de alta o baja calidad).

- a) Resume, visualiza y analiza los datos.
- b) Construye algunos clasificadores interesantes basados en SVM (explora diferentes kernels).

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 final de `repat4b.pdf`

Quizás de utilidad para la visualización:

R:

<http://www.sthda.com/english/wiki/visualize-correlation-matrix-using-correlogram>

Python:

<https://towardsdatascience.com/better-heatmaps-and-correlation-matrix-plots-in-python-41445d0f2bec>