

Control 1

1) [1.5 pts] *Cross-Validation*

Explica en tus palabras:

- ¿Para qué sirve el procedimiento de *Cross-Validation*?
- ¿Cómo funciona *Cross-Validation*?

2) [2 pts] Evaluación de desempeño

Considera un modelo m que decide si un video en internet es seguro para niños o no. El modelo al recibir como *input* un video y nos entrega un puntaje entre 0 y 10000. Mientras más alto este número, más seguro está el modelo de que el video es seguro. Si ahora necesitamos que el modelo responda **True** o **False** en el caso de que el video sea clasificado como **seguro** o **no seguro** respectivamente, tenemos que decidir a partir de qué puntaje se considera que un video es seguro.

Explica en tus palabras:

- ¿Qué pasa si el puntaje a partir que se considera como **seguro** un video es bajo (por ejemplo, 3000)?
- ¿Qué pasa si el puntaje a partir que se considera como **seguro** un video es alto (por ejemplo, 8000)?

Se espera que además incluya en cada caso cómo espera que se vea la matriz de confusión y qué pasa con la precisión y el *recall*.

3) [2.5 pts] Regresión polinomial

Considera una secuencia de puntos en el plano $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ que tú consideras que distribuyen según un polinomio de grado k . Para ajustar un polinomio de grado k :

$$\hat{y}(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k$$

a estos datos lo que hacemos es entrenar un modelo con el algoritmo de **regresión lineal multivariable**¹ en donde el *dataset* de entrada sería:

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & \dots & x_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^k \end{pmatrix}$$

- Explique por qué hacer esto efectivamente funciona.

Hint: piensa en cómo se ve el problema de optimización de los errores al cuadrado cuando queremos ajustar una regresión lineal de k variables. También considera que en ambos casos estás aprendiendo los coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_k$.

Detalles académicos

La respuesta a este control debe ser un **archivo .pdf** que vas a subir al buzón asociado a webcursos. La fecha límite es el viernes 02 de octubre a las 20:00 horas. Este control lo pueden hacer en grupos de hasta 2 personas. Recuerda que solo uno de los integrantes debe entregar el control y que el archivo debe tener el nombre de los integrantes.

¹Este algoritmo fue visto en la clase de regresión lineal.