

Instrucciones

- La tarea se debe subir a Aula el día Martes 30 de Junio a más tardar las 23:59.
- Explique claramente cada una de las conclusiones, incluyendo gráficos y referencias cuando sea necesario.

Pregunta 1 (30%): Regresión Lineal

1. Regresión lineal multiple-output (K. Murphy, Ejercicio 7.2). Cuando tenemos múltiples outputs independientes en la regresión lineal, el modelo es

$$p(y|x_i, W) = \prod_{j=1}^M \mathcal{N}(y_j, w_j^T x_i, \sigma_j^2) \quad (1)$$

Dado que la verosimilitud factoriza, también el MLE:

$$\hat{W} = [\hat{w}_1, \dots, \hat{w}_M] \quad (2)$$

$$\hat{w}_j = (X^T X)^{-1} Y_{:,j} \quad (3)$$

Suponer que tenemos los siguientes datos, donde x_i son datos binarios:

$$\begin{pmatrix} 0 & (-1, -1)^T, \\ 0 & (-1, -2)^T, \\ 0 & (-2, -1)^T, \\ 1 & (1, 1)^T, \\ 1 & (1, 2)^T, \\ 1 & (2, 1)^T \end{pmatrix} \quad (4)$$

Además transformaremos cada x_i utilizando la siguiente función base:

$$\phi(0) = (1, 0)^T, \phi(1) = (0, 1)^T \quad (5)$$

Siendo el modelo entonces

$$\hat{y} = W^T \phi(x) \quad (6)$$

Con W una matrix 2×2 . Calcular el MLE de W de los datos entregados.

Pregunta 2 (40%): Regresión Logística:

1. Derive el siguiente gradiente:

$$\frac{\partial L(\hat{y}, y)}{\partial h} = -\frac{\partial}{\partial h} \sum_i^N y_i \log(\hat{y}_i)$$

Donde $\hat{y} = \text{Softmax}(h)$ es la función Softmax vista en clases.

2. Implemente la regresión Logística en Python, para esto siga las instrucciones en el Notebook adjunto.

Pregunta 3 Naive Bayes (30%):

1. Análisis bayesiano de Naive Bayes. Considere el modelo Naive Bayes visto en clases:

$$p(x|y = c, \theta) = \prod_{j=1}^D p(x_j|y = c, \theta_{jc})$$

trataremos de implementar un modelo bayesiano para naive bayes. Para esto, se propone un a priori factorizado para los parámetros

$$p(\theta) = p(\pi) \prod_{j=1}^D \prod_{c=1}^C p(\theta_{jc})$$

donde π sigue una distribución de *Dirichlet* $D(\alpha)$ y cada θ_{jc} sigue una distribución *Beta*(1, 1)

- (a) Explique cual son las debilidades del modelo Naive Bayes original y porque es buena idea utilizar un enfoque bayesiano para el modelo Naive Bayes..
 - (b) Derive las distribuciones $p(\pi|D)$ y $p(\theta_{jc}|D)$
 - (c) Derive el posterior predictivo $p(y = c|x, D)$
2. Ajustando un modelo Naive Bayes a mano (Ejercicio 3.22 Murphy K.). Considere un modelo Naive Bayes (versión Bernoulli) para clasificación de spam con el vocabulario $V = \{\text{"secret", "offer", "low", "price", "valued", "customer", "today", "dollar", "million", "sports", "is", "for", "play", "healthy", "pizza"}\}$.

Tenemos los siguientes ejemplos de spam "million dollar offer", "secret offer today", "secret is secret" y de mensajes no-spam, "low price for valued customer", "play secret sports today", "sports is healthy", "low price pizza". Dado los estimadores de máxima verosimilitud, calcule lo siguiente: θ_{spam} , $\theta_{secret|spam}$, $\theta_{secret|no-spam}$, $\theta_{sports|no-spam}$, $\theta_{dollar|spam}$ donde $\theta_{j|c}$ es la probabilidad de que la característica j ocurra en la clase c .