Felipe Luque -201414222

Final Exam

1. True and False. Clearly and throughly justify your answer. (3.5 points)

(a) In classic econometrics maximizing the variance is the same as minimizing the MSE.

Falso

Se sabe que el MSE de un estimador $\hat{\theta}$ con respecto a un parámetro desconocido θ es:

$$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$$

$$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \mathrm{E}(\hat{\theta}) + \mathrm{E}(\hat{\theta}) - \theta)^2\right]$$

$$MSE(\hat{\theta}) = E\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)^2 + 2\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)\left(\mathrm{E}(\hat{\theta}) - \theta\right) + \left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right]$$

$$MSE(\hat{\theta}) = E\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)^2\right] + 2E\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)\left(\mathrm{E}(\hat{\theta}) - \theta\right)\right] + E\left[\left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right]$$

$$MSE(\hat{\theta}) = E\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)^2\right] + 2\left(\left(\mathrm{E}(\hat{\theta}) - \theta\right)\left(\mathrm{E}(\hat{\theta}) - \mathrm{E}(\hat{\theta})\right)\right) + E\left[\left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right]$$

$$MSE(\hat{\theta}) = E\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)^2\right] + E\left[\left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right]$$

Como se sabe que:

$$E\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)^2\right] = Var(\hat{\theta})$$

$$E\left[\left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right] = sesgo\ (\hat{\theta}, \theta)^2$$

Se tiene que el valor de MSE del parámetro es:

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + sesgo\ (\hat{\theta}, \theta)^2$$

A partir de esta ecuación se puede deducir que un incremento en la varianza se ve reflejado en un incremento del MSE

(c) Boosting is a technique to reduce variance based on trees by using bootstrapped samples from the original sample.

Falso.

Aunque Boosting sea una técnica que reduzca la varianza, no utiliza la misma estimación de las medias del método bagging, donde promedia las muestras para encontrar el predictor.

$$\widehat{f_{bag}} = \frac{1}{B}\sum_{b=1}^{B}\widehat{f^b}(x)$$

En el caso de boosting este realiza un promedio ponderado donde se asignan cierto tipo de pesos como por ejemplo en el método AddBoost

$$L = \sum_{i=1}^{n} w_i^{(m)} \exp\,(-\frac{1}{2}\alpha_m y_i G_m(x_i))$$

(d) The magic of Robbin's formula is that you don't need to know the prior to get a posterior.

Verdadero

Esto se demuestra en primer lugar cuando por regla de Bayes y suponiendo que se conoce la densidad $g(\theta)$ tenemos:

$$E\{\theta|x\} = \frac{\int_0^\infty \theta p_\theta(x) g(\theta) d\theta}{\int_0^\infty p_\theta(x) g(\theta) d\theta}$$

Además, se supone que $p_{\theta_k}(x)$ se distribuye Poisson con media y varianza θ_k

$$\Pr(x_k = x) = p_{\theta_k}(x) = \frac{e^{-\theta_k}\theta_k^x}{x!}$$

Dado que no se conoce la anterior de $g(\theta)$ se supone que

$$E\{\theta|x\} = \frac{\int_0^\infty \theta \frac{e^{-\theta}\theta^x}{x!} g(\theta) d\theta}{\int_0^\infty \frac{e^{-\theta}\theta^x}{x!} g(\theta) d\theta}$$

$$E\{\theta|x\} = \frac{(x+1)\int_0^\infty \frac{e^{-\theta}\theta^{x+1}}{(x+1)!} g(\theta) d\theta}{\int_0^\infty \frac{e^{-\theta}\theta^x}{x!} g(\theta) d\theta}$$

La densidad marginal de x, es:

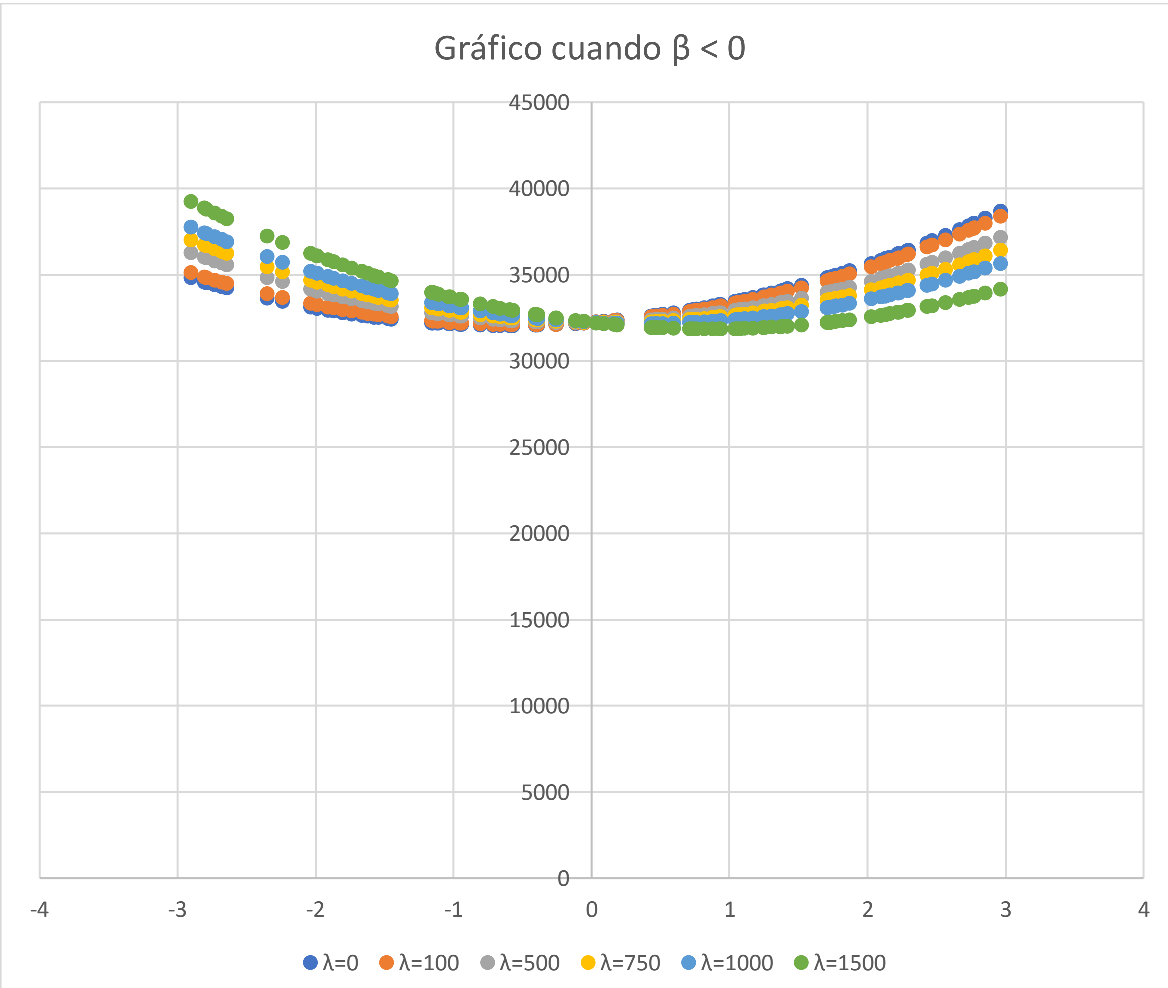
$$f(x) = \int_0^\infty p_{\theta_k}(x) g(\theta) d\theta = \int_0^\infty \frac{e^{-\theta}\theta^x}{x!} g(\theta) d\theta$$

Así tenemos que:

$$E\{\theta|x\} = \frac{(x+1)f(x+1)}{f(x)}$$

Esto demuestra que no es necesario conocer la prior de $g(\theta)$, para estimar la posterior.

2. In class I showed this plot. Replicate it for β < 0 and explain in detail what is going on under this new setting. (0.8 points)
   Utilizando un modelo en R y exportando los resultados, para distintos tipos de lambda se muestra la siguiente gráfica



A partir de los resultados se muestra simetría en los datos, aunque con un aumento considerable en el valor de lambda se inclina la curva hacia la derecha.

3. AdaBoost (0.7 points). In one of the last steps AdaBoost updates the following error function:

$$L = \sum_{i=1}^{n} w_i^{(m)} \exp\left(-\frac{1}{2}\alpha_m y_i G_m(x_i)\right)$$

show that if we differentiate that function with respect to α_m, then ADABoost are updating using:

$$\alpha_m = log\left(\frac{1 - err_m}{err_m}\right)$$

Se tiene la función:

$$L = \sum_{i=1}^{n} w_i^{(m)} \exp\left(-\frac{1}{2}\alpha_m y_i G_m(x_i)\right)$$

A partir de esta se pueden sacar dos funciones donde una representa la clasificación correcta de los datos de $G_m(x_i)$ y la otra para la clasificación equivocada de los datos de $G_m(x_i)$:

$$L = \sum_{i:G_m(x_i)=y_i}^{n} w_i^{(m)} \exp\left(-\frac{1}{2}\alpha_m\right) + \sum_{i:G_m(x_i)\neq y_i}^{n} w_i^{(m)} \exp\left(\frac{1}{2}\alpha_m\right)$$

Factorizando se tiene que

$$L = \left(\exp\left(\frac{1}{2}\alpha_m\right) - \exp\left(-\frac{1}{2}\alpha_m\right)\right)[\exp\left(-\frac{1}{2}\alpha_m\right)\sum_{i=1}^{n} w_i^{(m)} + \sum_{i=1}^{n} w_i^{(m)} I(G_m(x_i) \neq y_i)]$$

A partir de esto se deriva con respecto a α_m para encontrar el valor óptimo:

$$\frac{dL}{d\alpha_m} = \frac{1}{2}\alpha_m\left(\exp\left(\frac{1}{2}\alpha_m\right) - \exp\left(-\frac{1}{2}\alpha_m\right)\right)\left(-\frac{1}{2}\alpha_m\right)[\exp\left(-\frac{1}{2}\alpha_m\right)\sum_{i=1}^{n} w_i^{(m)} = 0$$

Parra simplificar la ecuación se divide por $\frac{\alpha_m}{2\sum_{i=1}^{n} w_i^{(m)}}$:

$$\exp\left(\frac{1}{2}\alpha_m\right)Err_m - \exp\left(-\frac{1}{2}\alpha_m\right)Err_m - \exp\left(-\frac{1}{2}\alpha_m\right) = 0$$

$$\exp\left(\frac{1}{2}\alpha_m\right)Err_m - \exp\left(-\frac{1}{2}\alpha_m\right)(1 - Err_m) = 0$$

$$\exp\left(\frac{1}{2}\alpha_m\right)Err_m = \exp\left(-\frac{1}{2}\alpha_m\right)(1 - Err_m)$$

Usando logaritmo en ambos lados:

$$\frac{1}{2}\alpha_m + \ln(Err_m) = -\frac{1}{2}\alpha_m + ln(1 - Err_m)$$

Así tenemos que α_m, es:

$$\alpha_m, = ln(1 - Err_m) - \ln(Err_m)$$

$$\alpha_m = ln\left(\frac{1 - Err_m}{Err_m}\right)$$