UNIVERSIDAD DE CONCEPCIÓN

FACULTAD DE INGENIERÍA

DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

Tarea n°3: Sistemas Adaptativos – 543820.

Jocelyn Matus Ancavil.

Profesor, Daniel Sbarbaro.

Concepción, lunes 15 de noviembre 2021

Problema 1

En este problema aplicaremos el enfoque Bayesiano para el estimar los parámetros de un modelo lineal en los parámetros. Consideremos que tenemos un conjunto con 40 mediciones igualmente espaciadas en el intervalo [-2 2] de la función (4.1) las cuales están contaminadas con ruido con distribución Gaussiana de media cero y desviación estándar y un modelo polinomial definido por (4.1)

gk(p) = x0 + x1p + x2 p2 +⋯+ xk pk

a) Derive la solución que minimiza el error cuadrado medio

$$\boldsymbol{F(x) = \sum_{q=1}^{Q} \left(t_k - g_k(P_q)\right)^2}$$

Se define

$$\mathrm{t} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} \quad \mathrm{G} = \begin{bmatrix} 1 & \cdots & {p_1}^k \\ \vdots & \ddots & \vdots \\ 1 & \cdots & {p_Q}^k \end{bmatrix} \quad \mathrm{x} = [x_0 \quad \cdots \quad x_n]$$

$$\boldsymbol{E_D = E_D = F(x) = \sum_{q=1}^{Q} \left(t_k - g_k(P_q)\right)^2} \qquad \textbf{(1)}$$

$$E_D = F(x) = [t - G\mathrm{x}]^T[t - G\mathrm{x}] \qquad \textbf{(2)}$$

$$F(x) = t^T t - 2\mathrm{x}^T G^T t + \mathrm{x}^T G^T X \qquad \textbf{(3)}$$

ara encontrar el mínimo, donde el gradiente de la función es cero

$$\boldsymbol{0 = \nabla F(x) = -2G^T t + 2\mathrm{x}^T G^T X} \qquad \textbf{(4)}$$

Entonces

$$G^T t = [G^T G] X^{Ml} \qquad \textbf{(5)}$$

Resolviendo, se obtiene la solución del mínimo cuadrado (o máximo probabilidad del ruido gaussiano)

$$X^{Ml} = [G^T G]^{-1} G^T t \qquad \textbf{(6)}$$

b) Derive la solución que minimice un indice regularizado

$$\boldsymbol{F(x) = \sum_{q=1}^{Q} \left(t_k - g_k(P_q)\right)^2 + \beta \sum_{i=1}^{k} (X_i)^2}$$

$$\boldsymbol{F(x) = \sum_{q=1}^{Q} \left(t_k - g_k(P_q)\right)^2 + \beta \sum_{i=1}^{k} (X_i)^2} \tag{7}$$

$$\mathbf{F(x)} = \boldsymbol{\alpha[t - G\mathbf{x}]^T[t - G\mathbf{x}] + \beta \mathbf{x}^T X} \tag{8}$$

$$\mathbf{F(x)} = \propto \boldsymbol{t^T t - 2 \propto \mathbf{x}^T G^T t + \propto \mathbf{x}^T G^T G X + \beta \mathbf{x}^T G^T X} \tag{9}$$

El mínimo se encuentra a través del gradiente de la función

$$\boldsymbol{0 = \nabla F(x) = -2 \propto G^T t + 2 \propto G^T G X + 2\beta \mathbf{X}} \tag{10}$$

$$0 = -2 \propto G^T t + 2 \propto G^T G X + \mathbf{2\beta X} \tag{11}$$

$$0 = -\alpha G^T t + 2\alpha G^T G X + \beta \mathbf{X} \tag{12}$$

$$\alpha G^T t = (\alpha G^T G + \beta) X \tag{13}$$

$$\mathrm{X} = (\alpha G^T G + \beta)^{-1} \alpha G^T t \tag{14}$$

c) Asuma que la densidad de probabilidad a priori es Gaussiana con una desviación standard σx y derive la expresión para la densidad de probabilidad condicionada $p(D|\ x, \alpha, \beta, k)$ sin el factor de normalización. Donde $\alpha=1/2\sigma_x^2$ y $\beta=1/2\sigma_\varepsilon^2$

$$p(x|\ D,\alpha,\beta,k)=\frac{p(D|\ x,\beta,k)\ p(x|\ \alpha\ ,k)}{p(D|\alpha,\beta,k)} \quad \textbf{(15)}$$

$p(D|\ \alpha,\beta,k)$ *es el factor de normalización.*
D el conjunto de datos
X el vector de parámetro
K es el orden del polinomio

entonces sin el factor de normalización

$$p(x|\ D,\alpha,\beta,k)=p(D|x,\beta,k)\ p(x|\ \alpha\ ,k) \quad \textbf{(16)}$$

Como se dice en el enunciado que la densidad de probabilidad a priori es Gaussiana con una desviación standard que solo depende de x, entonces

$$p(x)=\frac{1}{Z_w(\alpha)}\exp\left(-\beta E_w\right) \quad \textbf{(17)}$$

Donde

$Z_w(\beta)=(2\pi\sigma^2 w)^{N/2}=({\pi/\alpha})^{N/2}$

Entonces reemplazando 17 en 18, se tiene

$$p(D|\ x,\alpha,\beta,k)=p(D|x,\beta,k)\ \frac{1}{Z_w(\alpha)}\exp\left(-\alpha E_w\right) \quad \textbf{(18)}$$

De clases se tiene que

$$p(x|\ D,\alpha,\beta,k)=\frac{1}{z_f(\alpha,\beta,k)}\exp\left(-F(x)\right) \quad \textbf{(19)}$$

Despejando de 18

$$p(D|x,\beta,k)=\frac{1}{p(D|\ x,\alpha,\beta,k)}\frac{1}{Z_w(\alpha)}\exp\left(-\alpha E_w\right) \quad \textbf{(20)}$$

Reemplazando 19 en 20

$$p(D|x,\beta,k)=\frac{z_f(\alpha,\beta,k)}{\exp\left(-F(x)\right)}\frac{1}{Z_w(\alpha)}\exp\left(-\beta E_w\right)$$

d) Encuentre el valor de los parámetros (x_0, x_1, x_2, $\cdots$, x_k) que maximicen esta densidad de probabilidad.

Se tiene del ejercicio anterior que

$$p(D|x,\beta,k) = \frac{z_f(\alpha,\beta,k)}{\exp\big(-F(x)\big)}\frac{1}{Z_w(\alpha)}\exp\left(-\beta E_w\right) \quad \textbf{(21)}$$

Derivando esa expresión

$$\frac{\partial}{\partial x}\big(p(D|x,\alpha,\beta,k)\big) = \frac{\partial}{\partial x}\left(\frac{1}{z_f(\alpha,\beta)}e^{-F(x)}\cdot\frac{1}{z_x(\sigma_x)}e^{(-\beta E_w)}\right) \quad \textbf{(22)}$$

$$\frac{\partial}{\partial x}\big(p(D|x,\alpha,\beta,k)\big) = \frac{1}{z_f(\alpha,\beta)z_x(\sigma_x)}\left(\frac{\partial}{\partial x}\left(e^{-F(x)}\cdot e^{(-\beta E_w)}\right)\right) = 0 \quad \textbf{(23)}$$

Como $\frac{1}{z_f(\alpha,\beta)z_x(\sigma_x)}$ es un valor

$$\frac{\partial}{\partial x}\big(p(D|x,\alpha,\beta,k)\big) = \frac{\partial}{\partial x}\left(e^{-F(x)}\cdot e^{(-\beta E_w)}\right) = 0 \quad \textbf{(24)}$$

$$\frac{\partial}{\partial x}\big(p(D|x,\alpha,\beta,k)\big) = \frac{\partial}{\partial x}\left(e^{-\left(\alpha t^T t - 2\alpha x^T G^T t + \alpha x^T G^T G x + 2\beta x^T x\right)}\right) = 0 \quad \textbf{(25)}$$

$$0 = -(-2\alpha G^T t + 2\alpha G^T G x + 4\beta x)\left(e^{-\left(\alpha t^T t - 2\alpha x^T G^T t + \alpha x^T G^T G x + 2\beta x^T x\right)}\right) \quad \textbf{(26)}$$

Como la exponencial es una función de soporte infinito, es asíntota en cero entonces

$$0 = -(-2\alpha G^T t + 2\alpha G^T G x + 4\beta x) \quad \textbf{(27)}$$

Despejando x

$$x = [\alpha G^T G + 2\beta]^{-1}\alpha G^T t \quad \textbf{(28)}$$

Como $\alpha = 1/2\sigma_x^2$ y $\beta = 1/2\sigma_\varepsilon^2$

Entonces

$$x = \left[(1/2\sigma_x^2)G^T G + 2(1/2\sigma_\varepsilon^2)\right]^{-1}(1/2\sigma_x^2)G^T t \quad \textbf{(29)}$$

e) Derive la expresión para la densidad de probabilidad condicionada $p(x\ |D, \alpha, \beta, k)$ sin el factor de normalización y encuentre el valor de los parámetros que maximicen esta densidad de probabilidad

se tiene que

$$p(x|D,\alpha,\beta,k) = \frac{p(D|x,\beta,k)\cdot p(x|\alpha,k)}{p(D|\alpha,\beta,k)} \tag{30}$$

Sin el factor de normalización queda

$$p(x|D,\alpha,\beta,k) = p(D|x,\beta,k)\cdot p(x|\alpha,k) \tag{31}$$

Se tiene que

$$p(D|x,\beta,k) \tag{32}$$

Donde

X *el conjunto de datos*
D *el vector de parámetro*
K *es el orden del polinomio*
β *es el parámetro*

si se asume que tiene una distribución Gaussian y ruido constante, como se plantea en la clase

$$p(D|x,\beta,k) = \frac{1}{Z_D(\beta)} e^{(-\alpha E_D)} \tag{33}$$

Si x es el dato que se conoce antes de todo, ya que es la probabilidad a priori, entonces podemos suponer que se conoce su densidad asociada, ósea densidad a priori, si se supone que la densidad tiene distribución gaussiana, y x esta centrado en cero

$$p(x|\alpha,k) = \frac{1}{Z_w(\alpha)} e^{(-\beta E_w)} \tag{34}$$

Reemplazando 33 y 34 en 31 se tiene

$$p(x|D,\alpha,\beta,k) = \frac{1}{Z_D(\beta)} e^{(-\alpha E_D)} * \frac{1}{Z_w(\alpha)} e^{(-\beta E_w)} \tag{35}$$

$$p(x|D,\alpha,\beta,k) = \frac{1}{Z_D(\beta)} \frac{1}{Z_w(\alpha)} e^{(-(\alpha E_D + \beta E_w))} \tag{36}$$

Se define

$$F(x) = \alpha E_D + \beta E_w \tag{37}$$

Entonces reemplazando 37 en 36

$$p(x|D,\alpha,\beta,k) = \frac{1}{Z_D(\beta)}\frac{1}{Z_w(\alpha)}e^{(-(F(x)))} \quad \textbf{(38)}$$

Luego definiendo en Z_F

$$p(x|D,\alpha,\beta,k) = \frac{1}{Z_F(a,\beta)}e^{(-(F(x)))} \quad \textbf{(39)}$$

Donde

$$Z_F(a,\beta)\ depende\ solamente\ de\ a,\beta$$

Para que los valores de los parámetro maximicen la densidad, se debe minimizar F , como se vio en el ejemplo anterior

Entonces para encontrar el valor, se hace bajo criterio de newton

$$\frac{\partial}{\partial x}\left(p(x|D,\alpha,\beta,k)\right) = \frac{\partial}{\partial x}\left(\frac{1}{Z_F(\alpha,\beta)}e^{-F(x)}\right) = 0 \quad \textbf{(40)}$$

$$\frac{\partial}{\partial x}\left(\frac{1}{Z_F(\alpha,\beta)}e^{-F(x)}\right) = -(-2\alpha G^T t + 2\alpha G^T Gx + 2\beta x)\cdot e^{-(\alpha t^T t - 2\alpha x^T G^T t + \alpha x^T G^T Gx + 2\beta x^T x)} = 0 \quad \textbf{(41)}$$

Como la función exponencial es de soporte infinito, es asintótica en cero

$$-2\alpha G^T t + 2\alpha G^T Gx + 2\beta x = 0 \quad \textbf{(42)}$$

Entonces x

$$x = [\alpha G^T G + \beta]^{-1}\alpha G^T t \quad \textbf{(43)}$$

f) Derive la expresión para $p(D|\alpha, \beta, k)$ y comente porque la maximización de esta densidad de probabilidad es lo mismo que maximizar $p(\alpha, \beta\ |D, k)$ si se consideran que α y β tienen distribución de densidad uniforme.

Si

$$p(\alpha,\beta|D,k) = \frac{p(D|\alpha,\beta,k)p(\alpha,\beta,k)}{p(D|k)}$$

Como $p(D|k)$es el factor de normalización en este caso, no influye para este análisis y $p(\alpha,\beta,k)$ es un valor constante entonces se debe maximizar $p(D|\alpha,\beta,k)$, lo que ya fue hecho en los ejercicios anteriores

Entonces

$$p(D|\alpha,\beta,k) = \frac{p(D|x,\beta,k)\cdot p(x|\alpha,k)}{p(x|D,\alpha,\beta,k)}$$

puede se expresada sin considerar F(x) si se trabaja con el factor de normalización, como se mostrará a continuación

$$p(D|\alpha,\beta,k) = \frac{\left(\frac{1}{Z_D(\beta)}e^{(-\alpha E_D)}\right)\left(\frac{1}{Z_w(\alpha)}e^{(-\beta E_w)}\right)}{\frac{1}{Z_F(\alpha,\beta)}e^{-F(x)}} \tag{44}$$

$$p(D|\alpha,\beta,k) = \frac{\frac{1}{Z_D(\beta)Z_w(\alpha)}e^{(-F(x))}}{\frac{1}{Z_F(\alpha,\beta)}e^{-F(x)}} \tag{45}$$

$$p(D|\alpha,\beta,k) = \frac{Z_F(\alpha,\beta)}{Z_D(\beta)Z_w(\alpha)} \tag{46}$$

Para estimar $Z_F(\alpha,\beta)$ se va a considerar serie de Taylor usando la apromizacion del hessiano

$$F(x) \approx F(x^{MP}) + \frac{1}{2}(x - x^{MP})^T H^{MP}(x - x^{MP}) \tag{47}$$

Donde

$$H^{MP} = \beta\nabla^2 E_d + \alpha\nabla^2 E_W$$

$$p(x|D,\alpha,\beta,k) \approx \left(\frac{1}{Z_F(\alpha,\beta)}e^{-F(x^{MP})}\right)\left(e^{-\frac{1}{2}(x-x^{MP})^T H^{MP}(x-x^{MP})}\right) \tag{48}$$

Luego se tiene que una distribución de tipo Gaussian puede ser expresada como:

$$p(x) = \frac{1}{\sqrt{(2\pi)^k|(H^{MP})^{-1}|}}e^{\left(-\frac{1}{2}(x-x^{MP})^T H^{MP}(x-x^{MP})\right)} \tag{49}$$

Igualando estas últimas expresiones se tiene

$$\frac{1}{\sqrt{(2\pi)^k |(H^{MP})^{-1}|}} = \frac{e^{-F(x^{MP})}}{Z_F(\alpha,\beta)} \tag{50}$$

Luego una expresión aproximada

$$e^{-F(x^{MP})} * \sqrt{(2\pi)^k |(H^{MP})^{-1}|} = Z_F(\alpha,\beta) \tag{51}$$

Entonces

$$p(D|\alpha,\beta,k) = \frac{e^{-F(x^{MP})} * \sqrt{(2\pi)^k |(H^{MP})^{-1}|}}{Z_D(\beta) Z_w(\alpha)} \tag{52}$$

Como α y β tienen distribución de densidad uniforme. Se conoce $Z_D(\beta)Z_w(\alpha)$.

$$p(D|\alpha,\beta,k) = \frac{e^{-F(x^{MP})} * \sqrt{(2\pi)^k |(H^{MP})^{-1}|}}{\left(\frac{\pi}{\beta}\right)^{\frac{N}{2}} \left(\frac{\pi}{\alpha}\right)^{\frac{k}{2}}} \tag{53}$$

G) A partir de la maximización de p(D|α, β, k) encuentre las expresiones para α y β óptimos

Para encontrar el máximo, se aplica log a la función, para separar los términos

$$\ln\left(p(D|\alpha,\beta,k)\right) = ln\left(\frac{Z_F(\alpha,\beta)}{Z_D(\beta)Z_w(\alpha)}\right) \quad \textbf{(54)}$$

$$\ln\big(p(D|\alpha,\beta,k)\big) = \ln\big(Z_F(\alpha,\beta)\big) - \ln\left(Z_D(\beta)\right) + \ln\left(Z_w(\alpha)\right) \quad \textbf{(55)}$$

reemplazando

$$\ln\big(p(D|\alpha,\beta,k)\big) = \ln\big(e^{-F(x^{MP})} * \det(H^{MP})\big) - \ln\left(\frac{\pi}{\beta}\right)^{\frac{N}{2}} + \ln\left(\frac{\pi}{\alpha}\right)^{\frac{k}{2}} \quad \textbf{(56)}$$

$$\ln\big(p(D|\alpha,\beta,k)\big) = -F(x^{MP}) - \frac{1}{2}\ln(\det(H^{MP})) + \frac{N}{2}\ln(\beta) + \frac{k}{2}\ln(\alpha) + \frac{k}{2}\ln(2) - \frac{N}{2}\ln(\pi) \quad \textbf{(57)}$$

El sistema de ecuaciones queda

$$\frac{\partial}{\partial\alpha}\ln\big(p(D|\alpha,\beta,k)\big) = 0 \quad \textbf{(58)}$$

$$\frac{\partial}{\partial\beta}\ln\big(p(D|\alpha,\beta,k)\big) = 0 \quad \textbf{(59)}$$

Luego, la matriz hessiana es $H^{MP} = \beta\nabla^2 E_d + \alpha\nabla^2 E_W$ Entonces los valores propios de H^{MP}, si A= βE_d

$$\lambda_H = \lambda_A + 2\beta I \quad \textbf{(60)}$$

Se va a sumir que se puede expresas el determinante de la matriz hessiana como el producto de los valores propios de esta.

Entonces para α

$$\frac{\partial}{\partial\alpha}\left(\frac{1}{2}\ln\big(\det(H^{MP})\big)\right) = \frac{1}{2\det(H)}\frac{\partial}{\partial\alpha}\left(\prod_{i=1}^{k}\lambda_{A_i} + 2\beta I\right) = \sum_{i=1}^{k}\frac{1}{\lambda_{A_i} + 2\beta} = tr(H^{-1}) \quad \textbf{(61)}$$

Entonces para β

$$\frac{\partial}{\partial\beta}\left(\frac{1}{2}\ln\big(\det(H^{MP})\big)\right) = \frac{1}{2\det(H)}\frac{\partial}{\partial\beta}\left(\prod_{z=1}^{k}\lambda_{H_Z}\right) \quad \textbf{(62)}$$

$$\frac{\partial}{\partial\beta}\left(\frac{1}{2}\ln\left(\det\left(H^{MP}\right)\right)\right)= \frac{1}{2\det(H)}\frac{\partial}{\partial\beta}\left(\prod_{i=1}^{k}(\lambda_{A_i}+2\beta)\right) \tag{63}$$

se define como el numero efectivo de parámetros:

$$\gamma = k - 2\beta tr(H^{-1}) \tag{64}$$

$$\gamma = k - 2\beta\sum_{i=1}^{N}\frac{1}{\lambda_{A_i}+2\beta}=\sum_{i=1}^{k}\left(1-\frac{1}{\lambda_{A_i}+2\beta}\right)= \sum_{i=1}^{k}\left(\frac{\lambda_{A_i}}{\lambda_{A_i}+2\beta}\right) \tag{65}$$

$$\gamma = \sum_{i=1}^{k}\left(\frac{\lambda_{A_i}}{\lambda_{H_i}}\right) \tag{66}$$

Entonces

$$\frac{\partial}{\partial\beta}\frac{1}{2}\log(\det(H)) = \frac{1}{2\det(H)}\left(\sum_{i=1}^{k}\left(\prod_{j\neq i}\left((\lambda_{A_j}+2\beta\right)\right)\frac{\partial}{\partial\beta}\left((\lambda_{A_i}+2\beta\right)\right) \tag{67}$$

$$\frac{\partial}{\partial\beta}\frac{1}{2}\log(\det(H)) = \frac{1}{2}*\frac{\sum_{i=1}^{k}\left(\prod_{j\neq i}\left(\lambda_{A_j}+2\beta\right)\left(\frac{\lambda_{A_i}}{\alpha}\right)\right)}{\prod_{i=1}^{k}\left(\lambda_{A_i}+2\beta\right)} \tag{68}$$

$$\frac{\partial}{\partial\beta}\frac{1}{2}\log(\det(H)) = \frac{1}{2\alpha}\sum_{i=1}^{k}\frac{\lambda_{A_i}}{\lambda_{A_i}+2\beta} \quad = \frac{\gamma}{2\alpha} \tag{69}$$

Luego para la primera ecuación del sistema se tiene

$$\frac{\partial}{\partial\alpha}\log\left(p(D|\alpha,\beta,k)\right) = -\frac{\partial}{\partial\alpha}F(x^{MP})-\frac{1}{2}\frac{\partial}{\partial\alpha}\log(\det(H^{MP}))+\frac{\partial}{\partial\alpha}\frac{k}{2}\log(\alpha) = 0 \tag{70}$$

$$\frac{\partial}{\partial\alpha}\log\left(p(D|\alpha,\beta,k)\right) = -\frac{\partial}{\partial\alpha}\left(\alpha E_D(x^{MP})\right)-tr(H^{MP})^{-1}+\frac{k}{2\alpha^{MP}} = 0 \tag{71}$$

$$E_D(x^{MP}) = \frac{k}{2\alpha^{MP}} - tr(H^{MP})^{-1} \tag{72}$$

$$2\alpha^{MP}E_D(x^{MP}) = k - 2\alpha^{MP}tr(H^{MP})^{-1} \tag{73}$$

Despejando α y reemplazando $tr(H^{MP})^{-1}$

$$\alpha^{MP} = \frac{\beta^{MP}k}{2\beta^{MP}E_D(x^{MP})+k-\gamma} \tag{74}$$

Luego para la 2da ecuación del sistema de ecuaciones

$$\frac{\partial}{\partial\beta}\log\big(p(D|\alpha,\beta,k)\big) = -\frac{\partial}{\partial\beta}F(x^{MP}) - \frac{\partial}{\partial\beta}\frac{1}{2}\log(\det(H^{MP})) + \frac{\partial}{\partial\beta}\frac{N}{2}\log(\beta) = 0 \qquad \textbf{(75)}$$

$$\frac{\partial}{\partial\beta}\log\big(p(D|\alpha,\beta,k)\big) = -E_w(x^{MP}) - \frac{\gamma}{2\beta^{MP}} + \frac{N}{2\beta^{MP}} = 0 \qquad \textbf{(76)}$$

Despejando β^{MP}

$$\beta^{MP} = \frac{N-\gamma}{2E_w(x^{MP})} \qquad \textbf{(77)}$$

Finalmente, el sistema de ecuaciones queda

$$\alpha^{MP} = \frac{\beta^{MP}k}{2\beta^{MP}E_D(x^{MP}) + k - \gamma} \qquad \textbf{(78)}$$

$$\beta^{MP} = \frac{N-\gamma}{2E_w(x^{MP})} \qquad \textbf{(79)}$$

Reemplazando 79 en 78

$$\alpha^{MP} = \frac{(N-\gamma)k}{2E_D(x^{MP})(N-\gamma) + 2E_w(x^{MP})(k-\gamma)} \qquad \textbf{(80)}$$

Problema 2

a) Generar un conjunto de datos apropiados para realizar la modelación y comente

En este problema, en el cual se tiene los datos del consumo eléctrico de una casa, se va a elegir de salida la potencia activa medida, en kilowatt, de la casa.

Para poder seleccionar dos variables de entradas, se va a elegir entre los siguientes datos:

1) Global Intensity: Es la intensidad de corriente, medida en amperes, la cual es promediada por minuto, del gasto de la casa completa
2) Voltage: Es el voltaje usado por la casa, promediada en minutos, medida en volt.
3) Global reactive power: es la potencia reactiva, medida en kilowatt, promediada por minuto, de la casa completa
4) Sub metering 1, 2 y 3: Es el gasto de energía de potencia activa, en medidas de watts por hora, de varios sectores de la casa.

Para poder tener, de una manera objetiva y no gráficamente, una manera de poder buscar una correlación fuerte entre las variables independiente y la dependiente, se va a utilizar el algoritmo Relief-F, el cual enumera, desde el mayor al menor, cual variable esta más relacionada con la variable de salida.

Al ocupar este algoritmo, se encontró que las variables independientes más relacionadas con la salida son las siguientes: en primer lugar, la variable Global Intensity, en segundo lugar, la variable Voltage, en tercer lugar, seria Global reactive power, y en cuarto, quinto lugar y sexto lugar las variables Sub metering 2, Sub metering 1 y Sub metering 3, respectivamente.

Al obtener estos resultados, se eligen las dos primeras variables para variables independientes: Global Intensity y Voltage. Se puede ver en los siguientes gráficos como se relacionan estas variables con la variable de salida:

Relación entre la variable de entrada Global intensity y la variable de salida Global active power:

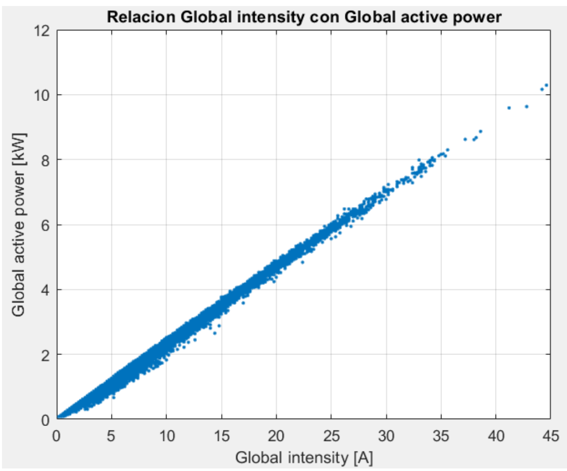


Figura N°1, Relación entre Global intensity y Global active power:

Relación entre la variable de entrada Voltage y la variable de salida Global active power:

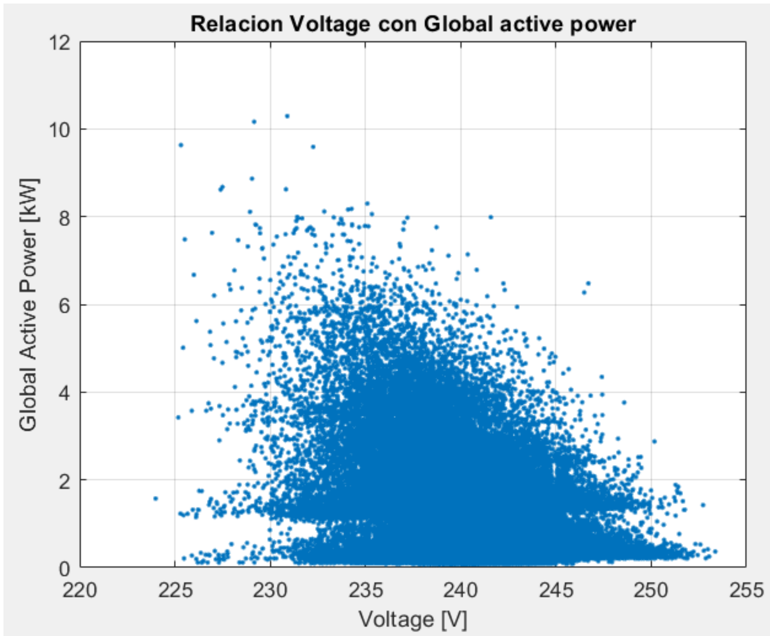


Figura N°2, Relación entre Voltage y Global active power

A partir de esto, para generar un conjunto apropiado para la modelación, ósea, entrenamiento, validación y testeo del modelo, se utilizó el comando *dividerand*, el cual divide, en ciertos porcentajes, los datos en 3 conjuntos: entrenamiento, validación y testeo, eligiendo los datos al azar

b) Realice los preprocesamientos necesarios de los datos.

Para el preprocesamiento de los datos, se va a realizar dos acciones:

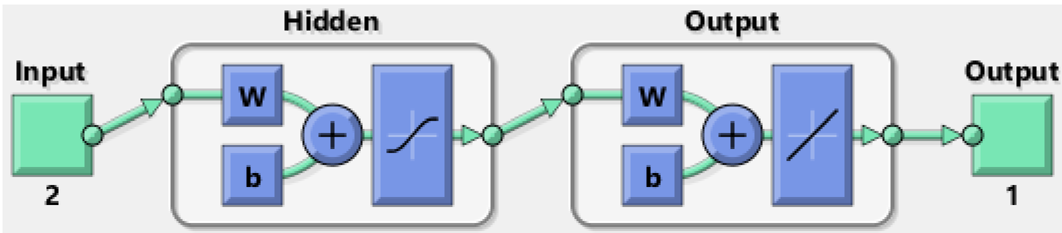
1) Se va a reducir la cantidad de datos que se va a manejar, ya que, para el computador utilizado, 2 millones de datos para cada variable, aproximadamente, es demasiado para que pueda trabajar con esta gran cantidad de datos.

Para reducir el tamaño se utiliza el comando *downsample*, el cual reduce la cantidad de datos en cierta cantidad de índice de datos. En este caso, el comando guarda el primer dato encontrado en el índice, y después guarda el dato encontrado en el índice 10, y así sucesivamente hasta el final, eliminando los datos que estén entre estos índices.

2) Para poder utilizar un entrenamiento correcto para las redes neuronales, se va a utilizar el comando *mapminmax*, el cual transforma los datos para que estos se encuentren entre -1 y 1.

c) Diseñe un modelo de dos capas que asegure una buena generalización. Comience con un modelo razonablemente grande y mediante regularización determine un modelo de menor complejidad. Vuelva ajustar los parámetros de ser necesario.

La red neuronal que se va a utilizar de aquí en adelante, sin considerar el tamaño de la capa oculta, es el siguiente:



La capa oculta tiene una función de transferencia de activación de tangente hiperbólica, mientras que la capa de salida tiene una función de activación lineal. Este modelo asegura que se tenga una buena generalización.

Para el caso de regularización, se utilizo el comando *feedforwardnet*, el cual, en sus opciones, se puede especificar si cual tipo de entrenamiento se quiere para la red neuronal. Para este caso, se eligió el entrenamiento de regularización bayesiana.

Para un modelo razonablemente grande, de 30 neuronas, se obtuvo los siguientes resultados:

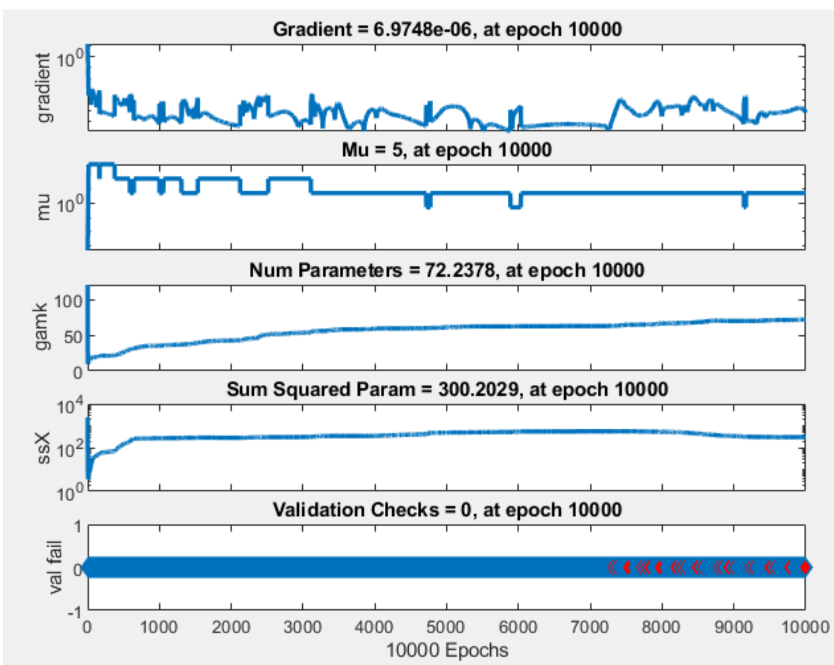


Figura N°3, Gráfico de los parámetros con 30 neuronas.

Al ser un modelo muy grande, tomó mucho tiempo en entrenarlo, por lo cual no es recomendable para que pueda ser entrenado de manera repetitiva. Además, al tener un gran tamaño, este puede realizar un *overfit* con nuevos datos, por lo que no es recomendable usar esta red.

Al utilizar distintos valores, se llegó a un tamaño óptimo de 8 neuronas para la capa oculta. Estos son los resultados que se obtuvieron:

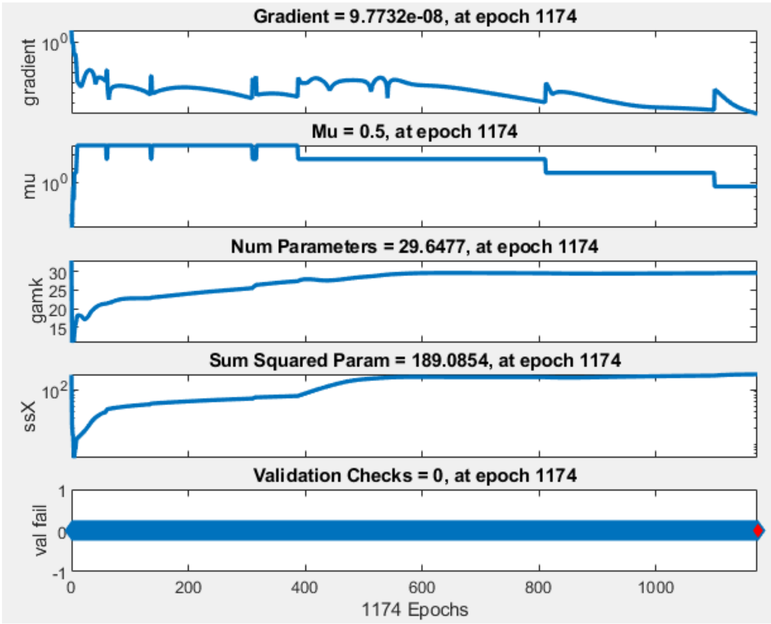


Figura N°4, **Gráfico de los parámetros con 8 neuronas**

En este modelo pudo converger hasta cierto punto, en el cual el gradiente llego a un valor tan bajo que, numéricamente, llego a un mínimo.

El número de parámetros efectivo pareciera ser alto, pero, aun así, con un rendimiento de 0.0000654 mse, es un muy buen rendimiento de la red neuronal.

d) Comience con un modelo de baja complejidad y usando validación cruzada determine la complejidad adecuada del modelo

Para el caso de entrenamiento de la red neuronal a través de una validación cruzada, se utilizo un *kfold* de 15, por lo cual se va a realizar 15 pasos para la validación cruzada.

Al usar una red neuronal de tamaño de 1 neurona de capa oculta, se obtuvo un rendimiento de 0.0000856 mse. Al cambiar el tamaño de la capa, se obtuvo una red neuronal de tamaño 5. El rendimiento de este fue de 0.0000767 mse. Al aumentar el tamaño de la red no hay mucha diferencia en el rendimiento, por lo que el tamaño de esta red es óptimo.

e) Cuantifique y analice la calidad de los modelos obtenidos por ambas estrategias

La calidad de ambos modelos es bastante buena, considerando la cantidad de datos y la complejidad de estas redes, al poder manejar gran cantidad de datos y tener un rendimiento muy bueno, con los valores que se mencionaron de rendimiento anteriormente.

La diferencia que se puede dar entre los dos modelos es que, a través de la regularización bayesiana, al aumentar el tamaño de la red, puede aumentar su rendimiento, pero esto no sucede con un entrenamiento de validación cruzada, el cual tiene un tope en su rendimiento.

La diferencia de tamaño de las dos redes no es tan importante, por lo que, en ese sentido, se puede elegir cualquiera de los dos.

Aun así, por su rendimiento, es mejor la red neuronal que fue entrenada a través de la regularización.

Gráficamente, se puede ver el rendimiento de estas redes en los siguientes gráficos:

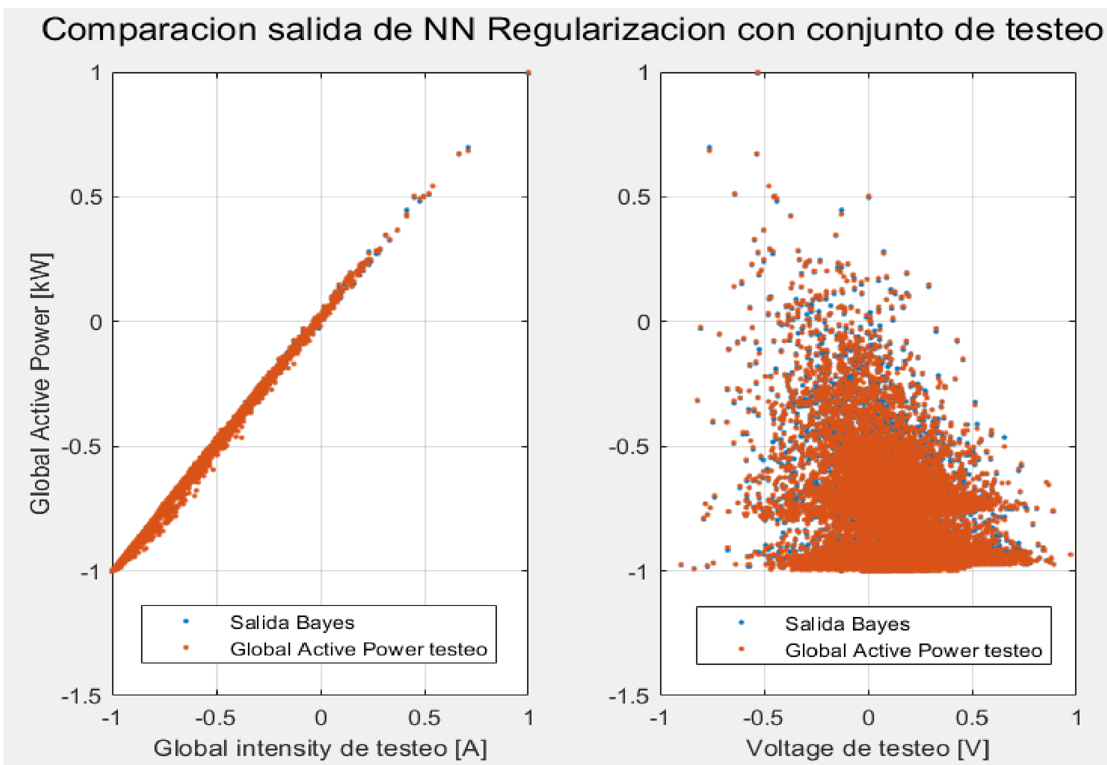


Figura N°5, Rendimiento de la red neuronal entrenada por regularización

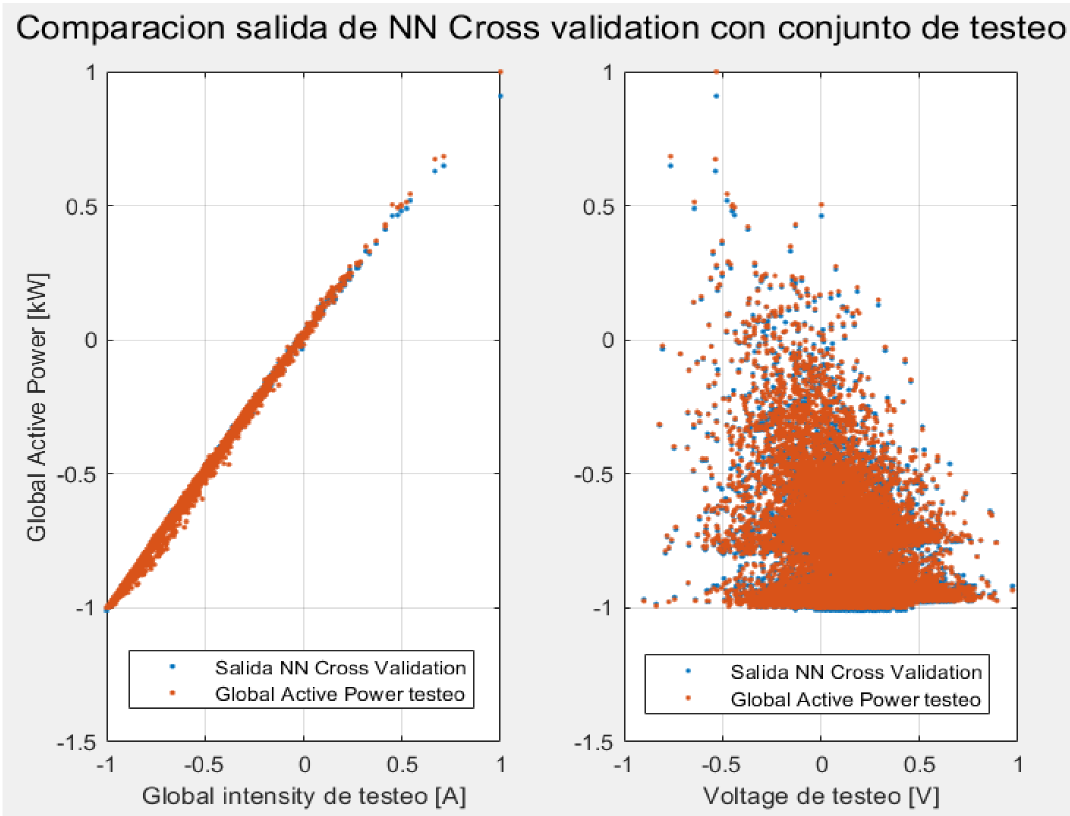


Figura N°6,   Rendimiento de la red neuronal entrenada a través de validación cruzada