Tarea 3: Introducción al Reconocimiento de Patrones

Jung Hwan Bak, Daniel Rojas Marín

18 Mayo, 2020

1 Primera Parte

1.1 Pregunta 1

1.1.1 Problema

Demuestre que para el caso general $J(\underline{\theta})$ se puede reescribir como $J(\underline{\theta}) = \frac{1}{2}(X\underline{\theta} - \underline{y})^T W(X\underline{\theta} - \underline{y})$ para una matriz diagonal W apropiada donde X es la matriz de diseño e $\underline{y}$ el vector de salida, tal y como lo definimos en clase.

1.1.2 Solución

$W_{ij} = 0$ siempre y cuando $i \neq j$ debido a que W es una matriz diagonal. Tomamos $W_i = w^i$ para todo $i = j$.
Sea $z = X\underline{\theta} - \underline{y}$ lo cual se traduce a $z_i = \underline{\theta}^T x^i - y^i$ entonces podemos obtener lo siguiente:

$$\tfrac{1}{2}(X\underline{\theta} - \underline{y})^T W(X\underline{\theta} - \underline{y}) = \tfrac{1}{2}\underline{z}^T W\underline{z}.$$

Toda matriz diagonal es simétrica, debido a esto podemos partir desde la suposición de que W es una matriz diagonal simétrica cuyos eigen-valores son reales, la expresión anterior se puede replantear como:

$$\tfrac{1}{2}\sum_{i=1}^{m} w^i z_i^2$$

Sustituimos a z_i^2 por su valor original:

$$\tfrac{1}{2}\sum_{i=1}^{m} w^i(\underline{\theta}^T x^i - \underline{y}^i)^2$$

Por lo tanto tenemos

$$J(\underline{\theta}) = \tfrac{1}{2}(X\underline{\theta} - \underline{y})^T W(X\underline{\theta} - \underline{y})$$

1.2 Pregunta 2

1.2.1 Problema

Encuentre la gradiente $\nabla_{\underline{\theta}} J(\underline{\theta})$ e iguálelo a cero para encontrar una versión generalizada de las ecuaciones normales en este contexto con ponderación. Encuentre una forma cerrada del valor de $\underline{\theta}$ que minimiza a $J(\underline{\theta})$, en función de X,W e $\underline{y}$.

1.2.2 Solución

Expansión de la función de error:

$$J(\underline{\theta}) = \tfrac{1}{2}(\underline{\theta}^T X^T - \underline{y})W(X\underline{\theta} - \underline{y})$$

$$= \tfrac{1}{2}(\underline{\theta}^T X^T W X\underline{\theta} - \underline{\theta}^T X^T W\underline{y} - \underline{y}^T W X\underline{\theta} + \underline{y}^T W\underline{y})$$

Que es un escalar y el último termino no depende de $\underline{\theta}$, por lo tanto:

$$\nabla_{\underline{\theta}} J(\underline{\theta}) = \nabla_{\underline{\theta}} \mathrm{tr}\{\tfrac{1}{2}(\underline{\theta}^T X^T W X\underline{\theta} - \underline{\theta}^T X^T W\underline{y} - \underline{y}^T W X\underline{\theta} + \underline{y}^T W\underline{y})\}$$

$$\nabla_{\underline{\theta}} J(\underline{\theta}) = \tfrac{1}{2}\nabla_{\underline{\theta}} \mathrm{tr}\{\underline{\theta}^T X^T W X\underline{\theta}\} - \tfrac{1}{2}\nabla_{\underline{\theta}} \mathrm{tr}\{\underline{\theta}^T X^T W\underline{y}\} - \tfrac{1}{2}\nabla_{\underline{\theta}} \mathrm{tr}\{\underline{y}^T W X\underline{\theta}\}$$

$$\nabla_{\underline{\theta}} J(\underline{\theta}) = X^T W X\underline{\theta} - X^T W\underline{y}$$

El mínimo se encuentra igualando a cero:

$$X^T W X\underline{\theta} - X^T W\underline{y} = 0$$

$$X^T W X\underline{\theta} = X^T W\underline{y}$$

$$\underline{\theta} = (X^T W X)^{-1} X^T W\underline{y}$$

1.3 Pregunta 3

1.3.1 Problema

Suponga que tenemos un conjunto de entrenamiento $\{(\underline{x}^{(i)}, y^{(i)}); i = 1...m\}$ de m ejemplos independientes, pero en el que los $y^{(i)}$ se observaron con varianzas distintas. Específicamente, suponga que

$$p(y^{(i)}|\underline{x}^{(i)};\underline{\theta}) = \tfrac{1}{\sqrt{2\pi}\sigma^{(i)}} \exp - \tfrac{(y^{(i)} - \underline{\theta}^T \underline{x}^{(i)})^2}{2(\sigma^{(i)})^2}$$

o en otras palabras, $y^{(i)}$ tiene media $\underline{\theta}^T \underline{x}^{(i)}$ y varianza $(^{(i)})^2$, donde las $\sigma^{(i)}$ son constantes fijas, conocidas. Demuestre que encontrar el estimado de máxima verosimilitud de $\underline{\theta}$ se reduce a resolver un problema de regresión lineal ponderada. Establezca claramente que los $w^{(i)}$ se calculan en términos de $\sigma^{(i)}$

1.3.2 Solución

Sabiendo que la verosimilitud se puede calcular con

$$L(\underline{\theta}) = \prod_{i=1}^{m} p(y^{(i)}|\underline{x}^{(i)};\underline{\theta})$$

y la verosimilitud logarítmica

$$l(\underline{\theta}) = lnL(\underline{\theta}) = ln\prod_{i=1}^{m} p(y^{(i)}|\underline{x}^{(i)};\underline{\theta})$$

$$= \sum_{i=1}^{m} lnp(y^{(i)}|\underline{x}^{(i)};\underline{\theta})$$

$$= \sum_{i=1}^{m} ln\frac{1}{\sqrt{2\pi}\sigma^{(i)}} \exp -\frac{(y^{(i)}-\underline{\theta}^T\underline{x}^{(i)})^2}{2(\sigma^{(i)})^2}$$

$$= \sum_{i=1}^{m} ln\frac{1}{\sqrt{2\pi}\sigma^{(i)}} + \sum_{i=1}^{m}(-\frac{(y^{(i)}-\underline{\theta}^T\underline{x}^{(i)})^2}{2(\sigma^{(i)})^2})$$

$$= \sum_{i=1}^{m} ln\frac{1}{\sqrt{2\pi}\sigma^{(i)}} - \frac{1}{2}\sum_{i=1}^{m}\frac{1}{(\sigma^{(i)})^2}(y^{(i)} - \underline{\theta}^T\underline{x}^{(i)})^2$$

El primer término es una constante, y por lo tanto maximizar la verosimilitud $l(\underline{\theta})$ se requiere minimizar el segundo término, que es idéntico a $J(\underline{\theta})$ haciendo

$$w^{(i)} = \frac{1}{(\sigma^{(i)})^2}$$

2 Segunda Parte

2.1 Pregunta 1

2.1.1 Problema

Use las ecuaciones normales para implementar la regresión lineal (no ponderada) $y = \underline{\theta}^T\underline{x}$ en el primer ejemplo de entrenamiento (esto es, la primera fila de train_qso). En una figura, grafique tanto los datos crudos como la línea recta resultante del ajuste. Indique el $\underline{\theta}$ óptimo resultante de la regresión lineal.

2.1.2 Solución

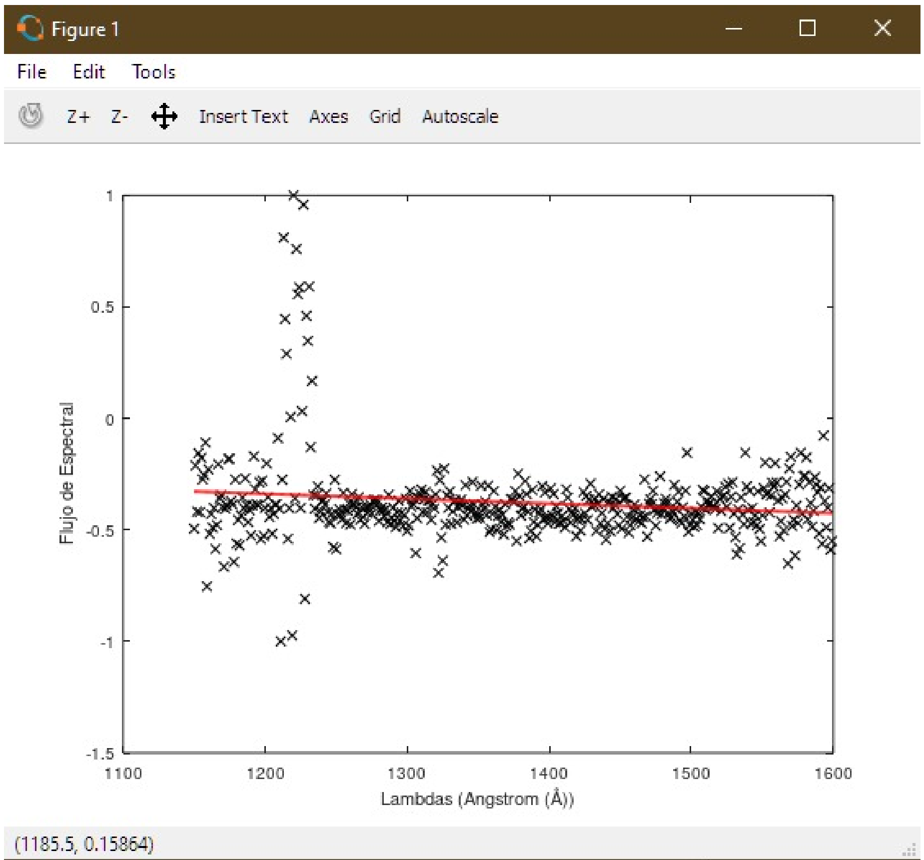

Figure 1: Solución a la Pregunta 1

2.2 Pregunta 2

2.2.1 Problema

Implemente la regresión lineal ponderada localmente en el primer ejemplo de entrenamiento. Use las ecuaciones normales que usted derivó en la pregunta 2 de la parte 1. En una figura aparte, grafique tanto los datos crudos como la curva suave resultante de su regresión. Cuando evalúe $h(\cdot)$ en un punto $\underline{x}$ use los pesos.

$$w^{(i)} = \exp\left(-\frac{||\underline{x}-\underline{x}^{(i)}||^2}{2\tau^2}\right)$$

con el parámetro de ancho de banda $\tau = 5$.

2.2.2 Solución

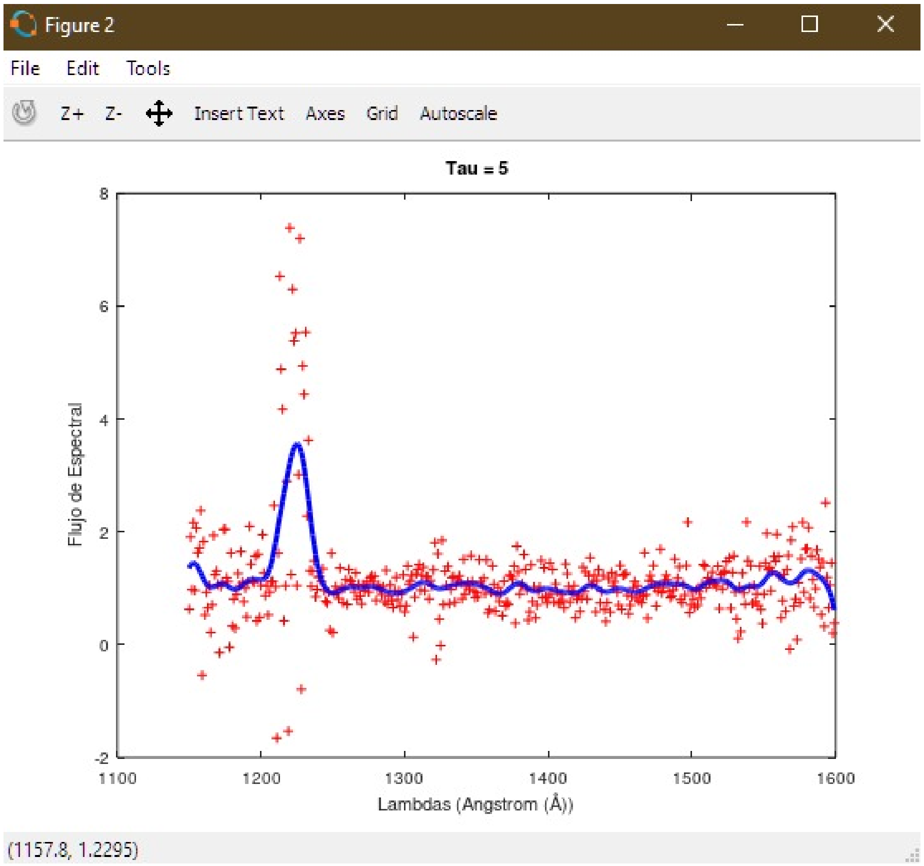

Figure 2: Solución a la Pregunta 2

2.3 Pregunta 3

2.3.1 Problema

Repita el problema 2 de la segunda parte cuatro veces más con $\tau = [1101001000]$. Grafique las curvas resultantes en una misma figura. Indique en una frase corta qué ocurre a la curva de regresión conforme τ crece.

2.3.2 Solución

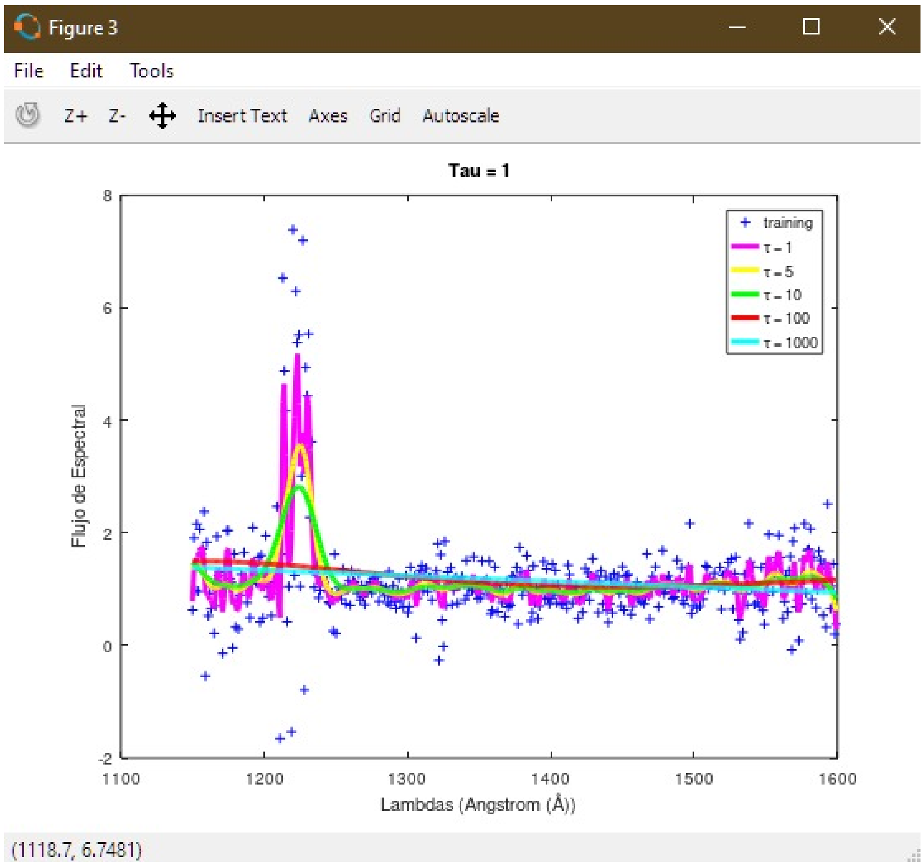

Figure 3: Solución a la Pregunta 3

Entre más abierta sea la campana gaussiana que se utiliza para ponderar los valores, más aplanada es la función generada en la regresión. Si la campana gaussiana es cerrada, la función presenta picos para ajustarse mejor a una sección más pequeña de valores.