Linear Regression with Laplace prior (on weights)

We assume the data to follow the model:

$$y_n = \vec{w}\,\vec{x}_n + \varepsilon_n, \quad \text{where } \varepsilon_n \sim \mathcal{N}(0,\sigma^2)$$

is noise (ε_n are i.i.d).

$\varepsilon = y - wx$

Prior is: $\vec{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_D \end{pmatrix}$, $\forall d$, $w_d \sim \frac{1}{2b} e^{-|w_d|/b}$.

$$w^*_{MAP} = \arg\max_{\vec{w}} \left(P(\vec{w} \mid X, Y) \right)$$

$$= \arg\max_{\vec{w}} \left(\frac{P(X,Y = \{x_n, y_n\}_n \mid \vec{w}) \cdot P(\vec{w})}{P(X,Y = \{x_n, y_n\}_n)} \right) \quad > 0$$

evidence: indep. from $\vec{w}$ $\rightarrow$ goes out

+(usual log trick)

$$= \arg\max_{\vec{w}} \left(\log \prod_{n=1}^{N} P((X_n,Y_n) = (x_n,y_n) \mid \vec{w}) + \prod_d \log\left(\frac{1}{2b} e^{-|w_d|/b}\right) \right)$$

const, ⊥ from $\vec{w}$

$$= \arg\max_{\vec{w}} \left(\sum_n^N \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\vec{w}\vec{x}_n - y)^2} \right) + \sum_d -\frac{|w_d|}{b} \right)$$

ε

indepdt from $\vec{w}$ $\Rightarrow$ out.

$$= \arg\max_{\vec{w}} \left(\sum_n^N -\frac{1}{2\sigma^2}(\vec{w}\vec{x}_n - y_n)^2 - \sum_d \frac{|w_d|}{b} \right)$$

$\times -\left(\frac{2\sigma^2}{N}\right)$

$$= \arg\min_{\vec{w}} \Bigg(\underbrace{\frac{1}{N}\sum_n^N (\vec{w}\vec{x}_n - y_n)^2 + \frac{2\sigma^2}{bN}\sum_d |\vec{w}_d|}_{=\mathcal{L}} \Bigg)$$

$\lambda \cdot |\vec{w}|_1$

This is now very similar to the 3.5. a.

For each w_d, $d = 1 \ldots D$, we have two cases: $w_d = 0$ or $w_d \neq 0$

$$\mathcal{L} = \sum_n^N \left(\sum_{d=1}^{D} w_d x_{nd} - y_n \right)^2 + \frac{2\sigma^2}{bN} \cdot \sum_{d=1}^{D} |w_d|$$

Case 1: $w_d \neq 0 \Rightarrow \nabla_{w_d}|w_d| = \text{sign}\, w_d$

$$\nabla_{w_d}\mathcal{L} = 0 \iff 0 = \nabla_{w_d}\left(\sum_n^N \Big(w_d x_{nd} + \underbrace{\sum_{d'\neq d} w_{d'} x_{nd'}}_{\perp\!\!\!\perp \text{ from } w_{d'}} - y_n\Big)^2 + \frac{2\sigma^2}{bN}\Big(|w_d| + \underbrace{\sum_{d'\neq d} |w_{d'}|}_{\perp\!\!\!\perp \text{ from } w_{d'}}\Big)\right)$$

$$\iff 0 = \left(\cancel{2}\sum_n^N \big(\underset{\text{(include } w_d \text{ also)}}{\vec{w}}\vec{x}_n - y_n\big)^1 \times x_{nd} + \frac{\cancel{2}\sigma^2}{bN}\,\text{sign}(w_d) + 0\right)$$

$$\iff 0 = \sum_n^N (w_d x_{nd} - y_n)\, x_{nd} + \underbrace{\Big(\sum_{d'\neq d}^D w_{d'} x_{nd'} - y_n\Big) x_{nd}}_{\text{let's call this } A_n} + \frac{\sigma^2}{bN}\,\text{sign}(w_d)$$

$$\sum_n - w_d x_{nd}^2 = \sum_n - y_n x_{nd} + \sum_n A_n + \frac{\sigma^2}{bN}\,\text{sign}(w_d)$$

let's assume $\sum_n^N x_{nd}^2 \neq 0$

$$w_d = \frac{1}{\sum_n^N x_{nd}^2}\left[\Big(\sum_n^N y_n x_{nd} - A_n\Big) - \frac{\sigma^2}{bN}\,\text{sign}(w_d)\right]$$

we divide by $\left(-\sum_n^N x_{nd}^2\right)$

$$= \underbrace{\frac{\sum_n^N (y_n x_{nd} - A_n)}{\sum_n^N x_{nd}^2}}_{\text{let's call this } B} - \underbrace{\frac{\frac{\sigma^2}{bN}}{\sum_{n=1}^N x_{nd}^2}}_{=C}\,\text{sign}(w_d)$$

(note: $C > 0$, obviously)

$$w_{d,MAP} = B - C \cdot \text{sign}(w_d)$$

B depends on the $w_{d'}$, $d' \neq d$, but not on w_d itself.

There are 2 sub-cases: $w_{d,MAP} > 0$, and < 0.

- $w_{d,MAP} > 0$: $B - C\,\text{sign}(w_d) > 0$
  $\iff B - C \times 1 > 0 \iff B > C \;(> 0)$
- $w_{d,MAP} < 0 \iff B - C \cdot \text{sign}(w_d) < 0$
  $\iff B + C < 0 \iff B < -C \;(< 0)$

In both cases, $|B| > C$.

In both cases, $\text{sign}(w_d) = \text{sign}(B)$

Case 2. $w_d = 0$: $\nabla_{w_d} |w_d| \in [-1, 1]$.

$\nabla_{w_d} \mathcal{L} = 0 \iff$ (similar to previous page)

$$\iff w_d = B - C \cdot \nabla_{w_d} |w_d| \quad , \text{ and } w_d = 0$$

$$B = C \cdot \nabla_{w_d} |w_d|$$

$$\nabla_{w_d} |w_d| = \frac{B}{C}$$ (assume $C \neq 0$, which is true, always)

We need to check that it's in $[-1, 1]$:

$$\nabla_{w_d} |w_d| \in [-1, 1] \iff -1 < B/C < 1$$

$$\iff -C < B < C$$

$$\iff |B| < C.$$

This is complementary to case 1, where $|B| > C$.

So, $w_{d,MAP}$ solution is summarized as:

(d, not d')

· Compute $B = \frac{1}{\sum_n x_{nd}^2} \left(\sum_n y_n x_{nd} - \sum_{d' \neq d} (w_{d'} x_{nd'} - y_n) x_{nd} \right)$

Compute $C = \frac{1}{\sum_{n=1}^{N} x_{nd}^2} \cdot \frac{\sigma^2}{bN}$

- If $|B| > C$, $w_d = B - C \operatorname{sign}(B)$ $\left(\operatorname{sign} B = \operatorname{sign} w_d\right)$
- If $|B| \leq C$, $w_d = 0$

At $|B| = C$, both solut° match.