Notes

Date / /

WiDS

ML from Scratch

W-1

3.1 $w = A^{-1}B$

$$\left[\begin{array}{cc|cc} 4 & 1 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{array}\right] \xrightarrow{R_2 - \frac{R_1}{4}} \left[\begin{array}{cc|cc} 4 & 1 & 1 & 0 \\ 0 & \frac{11}{4} & \frac{-1}{4} & 1 \end{array}\right]$$

$3 - \frac{1}{4}$

$\downarrow R_1 - \frac{4}{11}R_2$

$$\left[\begin{array}{cc|cc} 4 & 0 & \frac{12}{11} & \frac{-4}{11} \\ 0 & \frac{11}{4} & \frac{-1}{4} & 1 \end{array}\right]$$

$1 + \frac{\cancel{4}}{11} \times \frac{1}{\cancel{4}}$

$\downarrow \frac{R_1}{4},\ R_2 \times \frac{4}{11}$

$$\left[\begin{array}{cc|cc} 1 & 0 & \frac{3}{11} & \frac{-1}{11} \\ 0 & 1 & \frac{-1}{11} & \frac{4}{11} \end{array}\right]$$

$$w = \begin{bmatrix} \frac{3}{11} & \frac{-1}{11} \\ \frac{-1}{11} & \frac{4}{11} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{11} \\ \frac{7}{11} \end{bmatrix}$$

$$Aw = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{11} \\ \frac{7}{11} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = b$$

3.2

1. $\frac{df}{dx} = 2x + 3$

2. $\frac{\delta g}{\delta x_1} = 2x_1 + 2x_2$ $\quad \frac{\delta g}{\delta x_2} = 2x_1 + 6x_2$

3. $\frac{dh}{dx} = \begin{bmatrix} 2x \\ 3 \end{bmatrix}$

Notes Date / /

4. $\nabla_x g = \begin{bmatrix} \frac{\delta g}{\delta x_1} \\ \frac{\delta g}{\delta x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_1 + 6x_2 \end{bmatrix}$

5. $g(x_1, x_2) = x^T A x$

$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$g(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2$

$= [x_1 \quad x_2] \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= [ax_1 + cx_2 \quad bx_1 + dx_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$ax_1^2 + (b+c)x_1x_2 + dx_2^2$

$a = 1 \qquad d = 3 \qquad b + c = 2 \qquad b = c \Rightarrow = 1$

$A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$

6. $\frac{\delta}{\delta x}(x^T A x) = \frac{\delta}{\delta x}(g(x_1, x_2)) = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_1 + 6x_2 \end{bmatrix}$

$P \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_1 + 6x_2 \end{bmatrix}$

$\begin{bmatrix} 2 & 2 \\ 2 & 6 \end{bmatrix} = P = 2A$

Notes

Date / /

7. $\frac{\delta}{\delta x}(A^T x)$

(a) $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ column

$\frac{\delta}{\delta x}\left(\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} ax_1 + cx_2 \\ bx_1 + dx_2 \end{bmatrix}$

row vector

$A = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$

$A^T x = \sum_{i=1}^{n} a_i x_i$

A^T = row vector

$\frac{\delta}{\delta x_i}(A^T x) = a_i$

$A^T (a_1 \; a_2 \ldots a_n)$

$= \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} = A$

4

4.1 $P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|\bar{D})P(\bar{D})}$

$= \frac{0.99 \times 0.01}{\cancel{P(+|\bar{D})}\, 0.99 \times 0.01 + 0.05 \times 0.99}$

$P(\bar{D}) = 0.99$

$= \frac{99}{99 + 495} = \frac{99}{594} = 0.1666$

4.2

1. $L(\mu, \sigma^2) = p(x_1, x_2, \ldots x_n | \mu, \sigma^2) = \prod_{i=1}^{n} p(x_i | \mu, \sigma^2)$

Independent & identically distributed

Notes Date / /

$$p(x_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp \frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2$$

2. $l(\mu,\sigma^2) = \ln(L(\mu,\sigma^2))$

$$= \cancel{\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]\frac{1}{\sqrt{2\pi\sigma^2}}}$$

$$= n\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= n\ln(2\pi\sigma^2)^{-\frac{1}{2}} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

3. $\frac{\delta}{\delta\mu}(l) = \frac{-1}{2\sigma^2}\sum_{i=1}^{n} -2(x_i-\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$

$$= \frac{\sum_{i=1}^{n} x_i}{\sigma^2} - \frac{n\mu}{\sigma^2}$$

$$= 0$$

$$\frac{\sum_{i=1}^{n} x_i}{n} = \mu = \text{Mean}$$

Date / /

Notes

4. $\frac{\partial}{\partial(\sigma^2)}(l) = \frac{-n \times 2\pi}{2 \times 2\pi\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$

$= \frac{-n \times 2\pi}{4\pi\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$

$\frac{1}{2\sigma^{4^2}}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{n \times 2\pi}{4\pi\sigma^2} \Rightarrow \sigma^2 = \frac{+2\pi\sum_{i=1}^{n}(x_i-\mu)^2}{2\pi\, n}$

$= \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$

$=$ Variance

5. $\hat{\mu}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}$ $\quad \hat{\sigma}^2_{MLE} = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu}_{MLE})^2}{n}$

5.

1. 10

2. simple function $f(x) = 2x$

3. $8a + 4b + 2c + d = 4$ $\quad$ ~~$16a + 8b = 64a + 16b$~~

$64a + 16b + 4c + d = 8$ $\quad$ ~~$48a + 8b = 0$~~

$729a + 81b + 9c + d = 18$ $\quad$ ~~$-6a = b$~~

~~$4 = 48a + 6b +$~~ $56a + 12b + 2c = 4$ $\quad 28a + 6b + c = 2$

$665a + 65b + 5c = 10 \Rightarrow 133a + 13b + c = 2$

$28a - 90a + c = 2$ $\quad 28a + 6b + \cancel{c} = 133a + 13b + \cancel{c}$

$-62a + c = 2$ $\quad 7b + 105a = 0$

$b = -15a$

$133a - 195a + c = 2$ $\quad c = 2 + 62a$

Notes Date / /

$56a + 12(-15a) + 2(2+62a) = 4 \Rightarrow 56a + 124a - 180a + \cancel{4} = \cancel{4}$

$8a + 4(-15a) + 2(2+62a) + d = 4$

$(8 - 60 + 124)a + \cancel{4} + d = \cancel{4}$ $\quad 72a + d = 0$

$d = -72a$

$f(5) = a(125) - 15a(25) + (2+62a)5 - 72a$

$= (125 - 375 + 310 - 72)a + 10$

$= 435 - 447a + 10$

$= -12a + 10$

Need not be 10, only if $a = 0$

4. Assumption that the relationship is linear

5. The hypothesis used in linear regression is that is null hypothesis and its opposite, alternate hypothesis. If the model is

$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$

Null hypothesis isthe case when $\hat{\beta}_1 = 0$, implying no relation between x and y, and in alternate hypothesis, $\hat{\beta}_1 \neq 0$

If the standard error of $\hat{\beta}_1$ is less and $\hat{\beta}_1$ is large, then the null hypothesis can be

Notes Date / /

ignored. t-statistic $= \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$

~~t ↑ =1 Altern~~

p-value is the probability of how much the null hypothesis condition holds good. Lower p-value means relationship exists.

If p value is not very small, and greater than 1%, null hypothesis can be ~~rejected~~ confirmed.

6. If there are d features

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots \hat{\beta}_d x_d$$

For each x_i ($i \in [1, d]$), there is an associated ~~with~~ t-statistic and p-value.

If the p-value is sufficiently large for any particular x_i, we can confirm null hypothesis and remove that feature's contribution to y.

9. Considering data points $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$, loss for each term $= (y_i - \hat{y}_i)^2$ (squared)

~~total~~ $\hat{y} = wx + b$

$$L = \text{Total loss} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (wx_i + b)]^2$$

10. $$\frac{\delta L}{\delta w} = \sum_{i=1}^{n} 2(y_i - (wx_i + b))(-x_i)$$

$$= \sum_{i=1}^{n} 2x_i(wx_i + b - y_i) = 0$$

$$= \sum_{i=1}^{n} x_i(\cancel{wx_i + b - y_i})(y_i - (wx_i + b)) = 0$$

$$\frac{\delta L}{\delta b} = \sum 2(y_i - (wx_i + b))(-1) = 0$$

$$= \sum_{i=1}^{n} (y_i - (wx_i + b)) = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i - w\sum_{i=1}^{n} x_i - nb = 0$$

$$\sum y_i = w\sum x_i + nb$$

$$\frac{\sum y_i}{n} = \frac{w\sum x_i}{n} + b$$

$$\bar{y} = w\bar{x} + b$$

$$b = \bar{y} - w\bar{x}$$

Notes

Date / /

$$\cancel{\sum x_i y_i - w\sum x_i^2 + b\sum x_i = 0}$$

$$\cancel{\sum x_i y_i - w\sum x_i^2 + (\bar{y} - w\bar{x})\sum x_i = 0}$$

$$\sum x_i(y_i - w x_i - \bar{y} + w\bar{x}) = 0$$

$$= \sum x_i((y_i - \bar{y}) - w(x_i - \bar{x})) = 0$$

$$\sum x_i(y_i - \bar{y}) - w\sum x_i(x_i - \bar{x}) = 0$$

$$w = \frac{\sum x_i(y_i - \bar{y})}{\sum x_i(x_i - \bar{x})} = \frac{\sum x_i y_i - \bar{y}\sum x_i}{\sum x_i^2 - \bar{x}\sum x_i}$$

$$= \frac{\sum x_i y_i - \bar{y}\, n\bar{x}}{\sum x_i^2 - \bar{x}\, n\bar{x}}$$

$$= \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - \bar{x}^2 n}$$

$$\bar{x} = \frac{\sum x_i}{n} \Rightarrow \sum x_i = n\bar{x}$$

$$w = \cancel{\frac{n\bar{x}\sum(y_i - \bar{y})}{n\bar{x}\sum(x_i - \bar{x})}}$$

$$w = \frac{\sum x_i y_i - \bar{y}\sum x_i}{\sum x_i^2 - \bar{x}\sum x_i}$$

$$\sum(x_i - \bar{x})^2 = \sum(x_i^2 + \bar{x}^2 - 2x_i\bar{x})$$

$$= \sum x_i^2 - n\bar{x}^2 - 2\bar{x}\sum x_i$$

$$= \sum x_i^2 + n\bar{x}^2 - 2\bar{x}\, n\bar{x}$$

$$= \sum x_i^2 - n\bar{x}^2$$

$$\sum x_i(y_i - \bar{y}) = \sum x_i y_i - \bar{y}\sum x_i$$

$$= \sum x_i y_i - n\bar{x}\bar{y}$$

$\sum x_i y_i - n\bar{x}\bar{y}$

$\sum(x_i - \bar{x})(y_i - \bar{y})$

$= \sum x_i y_i$

$- \bar{x}\sum y_i$

$- \bar{y}\sum x_i$

$+ n\bar{x}\bar{y}$

Notes

Date / /

$$\frac{\sum x_i (y_i - \bar{y})}{\sum x_i (x_i - \bar{x})} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = w$$

11. $\hat{y} = wX$ $\quad w = \begin{bmatrix} b \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}$

12. $L = \|y - \hat{y}\|^2 = \|y - Xw\|^2$

$\|a\|^2$ $\quad a \Rightarrow$ column vector

$\|a\|^2 = a^T a \; [a_1 \, a_2 \ldots] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix}$ $\quad [a_1 = a_1^2 + a_2^2 + \ldots = \|a\|^2$

$L = (y - \cancel{wX})^T (y - \cancel{wX})$

13. $L = \cancel{(y - wX)}^T (y - Xw) = (y^T - X^T w^T)(y - Xw)$

$= \cancel{y^T y - X^T w^T y - y^T X w + X^T w^T X w}$

$= (y - w\cancel{X})^T (y - \cancel{w}X) = (y^T - w^T X^T)(y - wX)$

$= y^T y - w^T X^T y - y^T w X + w^T X^T w X$

Notes

Date / /

$[b \ w_1 \ \dots \ w_d] \begin{bmatrix} 1 & \dots & x_{1d} \\ \vdots & & \\ & & x_{nd} \end{bmatrix}$

$\to d$

$[\dots \ d] \begin{bmatrix} \vdots \\ d \end{bmatrix} = \text{Scalar}$

$(w^T x^T y)^T = \cancel{w x y^T} \ \cancel{y^T x} w \ y^T w x$

$L = y^T y - 2 w^T x^T y + w^T x^T w x$

$\frac{\delta L}{\delta w} = -2x^T y + 2x^T x w = 0$ (~~Max~~ Min L)

$2x^T y = 2x^T x w$

$\cancel{y} \ \cancel{w} = x^T y = x^T x w$

$w = (x^T x)^{-1} x^T y$