CS772 Machine Learning

Homework #2

Joseph S. Cannella

$(AB)C = A(BC)$
$A(B+C) = AB + AC$
$AB \neq BA$

CS-720 Homework #2

1.) $E(w) = \sum_{i=1}^{N} r_i (y_i - w^T x_i)^2$

$$E(\vec{w}) = (X\vec{w} - \vec{y})^T R (X\vec{w} - \vec{y})$$

$$= \vec{w}^T X^T R X \vec{w} - \underbrace{\vec{w}^T X^T R \vec{y} - \vec{y}^T R X \vec{w}}_{} + \vec{y}^T R \vec{y}$$

$(A^T)^T = A$; $(AB)^T = A^T B^T$;
$(A+B)^T = A^T + B^T$

Scalar

$$\underset{1 \times m}{[\vec{w}^T]} \underset{m \times N}{\begin{bmatrix} X^T \\ \\ \end{bmatrix}} \underset{N \times N}{\begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & \cdots & & r_N \end{bmatrix}} \underset{N \times 1}{\begin{bmatrix} y \\ \\ \end{bmatrix}} = 1 \times 1$$

m → # parameters ; N → # samples

$$E(\vec{w}) = \vec{w}^T X^T R X \vec{w} - 2 \cdot \vec{y}^T R X \vec{w} + \vec{y}^T R \vec{y}$$

$R = diag(r_1, \cdots, r_N)$

$$\frac{\delta E(\vec{w})}{\delta \vec{w}} = X^T R X \vec{w} - 2\vec{y}^T R X = \nabla E(\vec{w})$$

$$\min_w E(w) \rightarrow \nabla E(\vec{w}) = 0$$

$$X^T R X \vec{w} = 2\vec{y}^T R X$$

$$w = 2(X^T R X)^{-1} \vec{y}^T R X$$

$$\boxed{w_{min} = 2(X^T R X)^{-1} X^T R \vec{y}}$$

$$\frac{\delta(x^T a)}{\delta x} = \frac{\delta(a^T x)}{\delta x} = a^T$$

2) Let $X =$

$$X = \begin{bmatrix} 1 & 15.6 \\ 1 & 26.8 \\ 1 & 37.8 \\ 1 & 36.4 \\ 1 & 35.5 \\ 1 & 18.6 \\ 1 & 15.3 \\ 1 & 7.9 \\ 1 & 0 \end{bmatrix} ; \quad \vec{Y} = \begin{bmatrix} 5.2 \\ 6.1 \\ 8.7 \\ 8.5 \\ 8.8 \\ 4.9 \\ 4.5 \\ 2.5 \\ 1.1 \end{bmatrix} ; \quad \vec{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$X\vec{w} = \vec{Y} \quad ; \quad E(\vec{w}) = \sum_{i=1}^{N} (Y_i - x_i^T \vec{w})^2$$

$$E(\vec{w}) = \overbrace{(\vec{Y} - X\vec{w})^T}^{a} (\vec{Y} - X\vec{w})$$

$$\frac{\delta a}{\delta w} = \frac{\delta}{\delta w}(Y^T - X^T w^T$$

$$\frac{\delta E(\vec{w})}{\delta \vec{w}} = (\vec{Y} - X\vec{w})^T X + X^T(\vec{Y} - X\vec{w})$$

$$= 2 \cdot X^T(\vec{Y} - X\vec{w})$$

$$\min_{w} \nabla E(\vec{w}) = \cancel{2} X^T(\vec{Y} - X\vec{w}) = 0$$

$$X^T\vec{Y} = X^T X \vec{W}$$

$$\vec{w} = (X^T X)^{-1}(X^T \vec{Y})$$

Cont'd

$$\vec{w} = (X^T X)^{-1}(X^T \vec{y})$$

$$\underbrace{X^T X}_{2\times2} = \overset{2\times9}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 15.6 & 26.8 & 37.8 & 36.4 & 35.5 & 18.6 & 15.3 & 7.9 & 0 \end{bmatrix}} \overset{9\times2}{\begin{bmatrix} 1 & 15.6 \\ 1 & 26.8 \\ 1 & 37.8 \\ 1 & 36.4 \\ 1 & 35.5 \\ 1 & 18.6 \\ 1 & 15.3 \\ 1 & 7.9 \\ 1 & 0 \end{bmatrix}}$$

$$X^T X = \begin{bmatrix} 9 & 193.9 \\ 193.9 & 5618.11 \end{bmatrix}$$

$$\det(X^T X) = 9 \cdot 5618.11 - 193.9 \cdot 193.9 = 12{,}965.78$$

$$\operatorname{adj}(X^T X) = \begin{bmatrix} 5618.11 & -193.9 \\ -193.9 & 9 \end{bmatrix}$$

$$(X^T X)^{-1} = \operatorname{adj}(X^T X) \cdot \frac{1}{\det(X^T X)} = \begin{bmatrix} 0.4333 & -0.01495 \\ -0.01495 & 0.000694 \end{bmatrix}$$

$$X^T Y = \overset{2\times9}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 15.6 & 26.8 & 37.8 & 36.4 & 35.5 & 18.6 & 15.3 & 7.9 & 0 \end{bmatrix}} \overset{9\times1}{\begin{bmatrix} 5.2 \\ 6.1 \\ 8.7 \\ 8.5 \\ 8.8 \\ 4.9 \\ 4.5 \\ 2.5 \\ 1.1 \end{bmatrix}}$$

☆ Add Plot

$$X^T Y = \begin{bmatrix} 54.8 \\ 1375 \end{bmatrix}$$

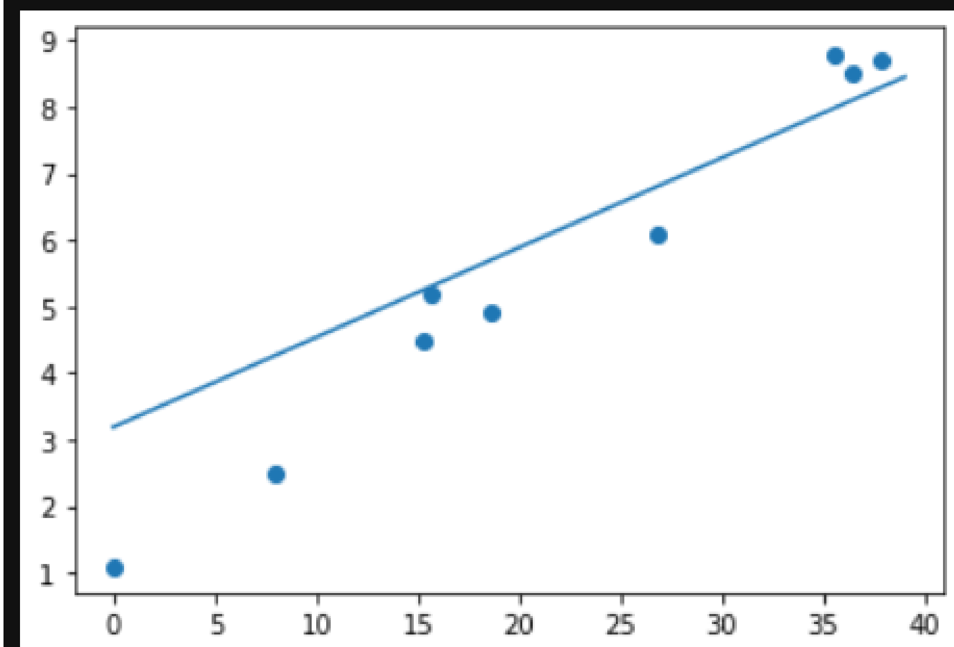
$$\vec{w}_{min} = \overset{2\times2}{\begin{bmatrix} 0.4333 & -0.01495 \\ -0.01495 & 0.000694 \end{bmatrix}} \overset{2\times1}{\begin{bmatrix} 54.8 \\ 1375 \end{bmatrix}} = \begin{bmatrix} 3.18859 \\ 0.13499 \end{bmatrix}$$

$$\boxed{\hat{Y}(x) = 3.18859 + x \cdot 0.13499}$$

```
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline
x = [15.6, 26.8, 37.8, 36.4, 35.5, 18.6, 15.3, 7.9, 0]
y = [5.2, 6.1, 8.7, 8.5, 8.8, 4.9, 4.5, 2.5, 1.1]
x_pred = np.arange(0.0, 40.0)
y_pred = list(map(lambda X: 3.188 + X*0.135, x_pred))

plt.scatter(x, y)
plt.plot(x_pred, y_pred)
```

```
[<matplotlib.lines.Line2D at 0x1cfca84ce48>]
```



$\frac{d(X^T A X)}{dX} = X^T A + X^T A^T$
$= X^T(A+A^T)$ $\left[\begin{array}{l} ex_2 \rightarrow \\ ex_3 \rightarrow \end{array}\right]$

$X^T A = A^T X$

$\frac{d(X^T A)}{dX} = \frac{d(a^T x)}{dx} = a^T$

Homework #2

3.) Given: X + $\bar{x} = \vec{0}$ $y = w_0 + wx$

$$E(w, w_0) = (\vec{y} - X\vec{w} - \vec{w}_0 1)^T(\vec{y} - X\vec{w} - \vec{w}_0 1) + \lambda \vec{w}^T\vec{w}$$

Proove: $E_{min}\begin{cases} \hat{w} = (X^T X + \lambda I)^{-1} X^T \vec{y} \\ \vec{w}_0 = \bar{y} \end{cases}$

$$E(\vec{w}, \vec{w}_0) = \vec{y}^T\vec{y} - \vec{y}^T X\vec{w} - \vec{y}^T\vec{w}_0 1 - \vec{w}^T X^T\vec{y} + \vec{w}^T X^T X\vec{w} + \vec{w}^T X^T \vec{w}_0 1 \dots$$
$$- 1^T\vec{w}_0^T\vec{y} + 1^T\vec{w}_0^T X\vec{w} + 1^T\vec{w}_0^T\vec{w}_0 1 + \lambda\vec{w}^T\vec{w}$$

$$= \vec{y}^T\vec{y} - 2\vec{y}^T X\vec{w} - 2\underset{a\times b}{\vec{y}^T}\underset{b\times b}{\vec{w}_0}\underset{b\times d}{1} + 2\cdot\vec{w}^T X^T\vec{w}_0 1 + 1^T\vec{w}_0^T\vec{w}_0 1 + \vec{w}^T X^T X\vec{w}$$
$$+ \lambda\vec{w}^T w$$

$$\frac{\partial E}{\partial w_0} = -2\vec{y}1 + 2\cdot\vec{w}^T X^T 1 + 2\,1^T\vec{w}_0^T 1$$

$$\frac{\partial E}{\partial w_0}\Big|_{min} = 0 = -\vec{y} + \vec{w}^T X^T + 1^T\vec{w}_0^T$$

$$\vec{y} = \vec{w}^T X^T + \vec{w}_0 1$$

$$\underbrace{\vec{y}(1)^{-1}}_{\bar{y}} = \vec{w}^T \underbrace{\cancel{X^T(1)^{-1}}}_{\bar{x}}{}^{0} + \vec{w}_0$$

$$\boxed{\vec{w}_0 = \bar{y}}$$

Cont'd →

$X^TA = A^TX$

$(AB)^T = B^TA^T$

4

Homework #2

3.) Cont'd

$$E(\vec{w},\vec{w}_0) = \vec{y}^T\vec{y} - \overbrace{2\vec{y}^TX\vec{w}}^{a} - 2\vec{y}^T\vec{w}_0 1 + \overbrace{2\cdot\vec{w}^TX^T\vec{w}_0 1}^{b} + \cdots$$

$$+ 1^T\vec{w}_0^T\vec{w}_0 1 + \overbrace{\vec{w}^TX^TX\vec{w}}^{c} + \overbrace{\lambda\vec{w}^T\vec{w}}^{d}$$

$$b = 2\vec{w}^TX^T\vec{w}_0 1 = 2\cdot(\vec{w}_0 1)^T(\vec{w}^TX)^T = 2\cdot 1^T\vec{w}_0^TX^T\vec{w}$$

$$\frac{\delta(X^TAX)}{\delta X} = X^T(A+A^T)$$

$$\delta b/\delta\vec{w} = 2\cdot 1^T\vec{w}_0^TX^T$$

$$\frac{\delta E}{\delta\vec{w}} = -2\vec{y}^TX + 2\cdot 1^T\vec{w}_0^TX^T + 2\vec{w}^TX^TX + 2\lambda\vec{w}^T$$

$$\delta E/\delta\vec{w}_{min} = 0$$

$$\vec{y}^TX - \cancel{1^T\vec{w}_0^TX^T} = \hat{w}^TX^TX + \lambda\hat{w}^T = \hat{w}(X^TX+\lambda I)$$

$$X^T\vec{y} = \hat{w}(X^TX+\lambda I)$$

$$\boxed{\hat{w} = (X^TX+\lambda I)^{-1}X^T\vec{y}}$$

Homework#2

4.) Linear Regression

$$\vec{Y} \approx f(X; \vec{w}_0, \vec{w}^T) = \vec{w}_0 + \vec{w}^T X$$

$$[w_0 w^T]^T = ([1X]^T [1X])^{-1} [1X]^T Y$$

Can solve $\hat{w}, \hat{w}_0$ seperately

$$E(X; \vec{w}_0, \vec{w}) =$$