Linear Regression in Matrix form

feature vector

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & & & \\ x_{m1} & x_{m2} & \cdots & x_{md} \end{bmatrix}$$

Observations

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

we want to find $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}$ that minimizes error of $Xw - Y$ for each data point

Error $e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix} = \begin{bmatrix} \hat{y}_1 - y_1 \\ \hat{y}_2 - y_2 \\ \vdots \\ \hat{y}_m - y_m \end{bmatrix}$

1 case | Matrix case

loss

$$\hat{L}(w) = \frac{1}{m} \sum_{i=1}^{m} \frac{(wx_i - y_i)^2}{2}$$

$$\hat{L}(w) = \frac{1}{m} \| Xw - y \|_2^2$$

gradient

$$\hat{L}'(w) = \frac{1}{m} \sum_{i=1}^{m} (wx_i - y_i) x_i$$

$$\nabla L(w) = \frac{1}{m} 2 X^T (Xw - y)$$

solution for w

$$w = \frac{\sum_{i=i}^{m} y_i x_i}{\sum_{i=1}^{m} x_i^2}$$

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

(1)

To find the loss

$$\frac{1}{m}\sum_{i=1}^{m} \ell(\hat{y}_i, y_i) = \frac{1}{m}\sum_{i=1}^{m} (\hat{y}_i - y_i)^2 = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=1}^{d} w_j x_{ij} - y_i\right)^2$$

$$\frac{1}{m} e^T e = \frac{1}{m}(Xw - y)^T (Xw - y)$$

dot product is like sum of components squared

$$= \underbrace{\frac{1}{m}\sum_{i=1}^{m}\left(\underbrace{\sum_{j=1}^{d} w_j x_{ij}}_{(Xw)_i} - y_i\right)^2}_{\text{2-norm squared}}$$

$$= \frac{1}{m}\|Xw - y\|_2^2$$

② Derive $\nabla L(w)$

Some useful things to know from matrix calc

Matrix derivatives

Let A be a $k \times k$ matrix of constants, a be a $k \times 1$ vector of constants, and y be a $k \times 1$ vector of variables

① If $z = a^T y$ then

$$\frac{\partial z}{\partial y} = \frac{\partial a^T y}{\partial y} = a$$

② If $z = y^T y$ then

$$\frac{\partial z}{\partial y} = \frac{\partial y^T y}{\partial y} = 2y$$

③ If $z = a^T A y$ then

$$\frac{\partial z}{\partial y} = \frac{\partial a^T A y}{\partial y} = A^T a$$

④ If $z = y^T A y$, then

$$\frac{\partial z}{\partial y} = \frac{\partial y^T A y}{\partial y} = Ay + A^T y$$

If A is symmetric, then

$$\frac{\partial y^T A y}{\partial y} = 2Ay$$

Derive $\nabla L(w)$

First play around with $L(w)$ so we can get a form that the matrix derivatives can be applied to

$$\frac{1}{m}(Xw-y)^T(Xw-y)$$

$$= \frac{1}{m}(w^T X^T - y^T)(Xw-y)$$

$$= \frac{1}{m}\left[w^T X^T (Xw-y) - y^T(Xw-y)\right]$$

$$= \frac{1}{m}\left[w^T X^T X w - \underbrace{\overset{\text{scalar}}{w^T X^T y}}_{=(y^T X w)^T} - \overset{\text{scalar}}{y^T X w} + y^T y\right]$$

$$= \frac{1}{m}\left[\underbrace{w^T X^T X w}_{\text{apply } 4} - \underbrace{2 y^T X w}_{\text{apply } 2} + \underbrace{y^T y}_{=0}\right]$$

constant

$$\nabla L(w) = \frac{1}{m}\left[X^T X w + X X^T w - 2y^T X \right]$$

equal

$$= \frac{1}{m}\left[2X^T X w - 2X^T y \right]$$

$$= \frac{1}{m} 2 X^T (Xw - y)$$

Now set to zero:

$$\frac{1}{m} 2X^T (Xw - y) = 0$$

$$\Rightarrow X^T X w - X^T y = 0$$

$$\Rightarrow X^T X w = X^T y$$

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

(1) (One variable quadratic regression) Consider (EL) with $X = \{1, 2, 3\}$, $Y = \{2, 4, 5\}$. Use linear model $h = w * x$. Solve the problem with the quadratic loss. Hint: write down $\hat{L}(w)$ and find the solution by solve $\hat{L}'(w) = 0$.

$$X = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad Y = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

linear model $h = w * x$

$h_w(x_i)$

Recall from notes that $\hat{L}(w) = \frac{1}{m} \sum_{i=1}^{m} \ell(w x_i, y_i)$

$$\hat{L}'(w) = \frac{1}{m} \sum_{i=1}^{m} \ell'(w x_i, y_i)$$

Quadratic loss $\ell(w x_i, y_i) = \frac{(w x_i - y_i)^2}{2}$

$$\hat{L}(w) = \frac{1}{3}\left(\frac{(w-2)^2}{2} + \frac{(2w-4)^2}{2} + \frac{(3w-5)^2}{2} \right)$$

$$\begin{aligned} \hat{L}'(w) &= \frac{1}{3}\left((w-2) + 2(2w-4) + (3w-5)^2 \right) \\ &= 1/3 \left(w - 2 + 4w - 8 + 9w - 15 \right) \\ &= 1/3 \left(14w - 25 \right) \\ &= 0 \end{aligned}$$

$$\Rightarrow \quad 14w = 25$$

$$\Rightarrow \quad \boxed{w = 25/14}$$

We can also solve this more straightforwardly by considering the formula derived in class for the 1D case:

Since $\ell(wx_i, y_i) = \frac{(wx_i - y_i)^2}{2}$

$$\ell'(wx_i, y_i) = (wx_i - y_i)\, x_i$$

$$\frac{1}{m}\sum_{i=1}^{m} \ell'(wx_i, y_i) = \frac{1}{m}\sum_{i=1}^{m} (wx_i - y_i)\, x_i = 0$$

Solving gives us $w = \dfrac{\sum_{i=1}^{m} y_i x_i}{\sum_{i=1}^{m} x_i^2}$

We can plug in: $w = \dfrac{2 \cdot 1 + 4\ 2 + 5 \cdot 3}{1^2 + 2^2 + 3^2} = \boxed{\dfrac{25}{14}}$

(2) (One variable ℓ_1 regression) Same setup as in the previous problem: consider (EL) with $X = \{1, 2\}$ $Y = \{1, 3\}$. Use linear model $h = w * x$. Solve the problem with the ℓ_1 loss. Hint: plot the function $\hat{L}(w)$, which is piecewise linear, and find the minimum value (by finding the intersection of two lines).

$$X = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad Y = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

linear model $h = w * x$

ℓ_1 loss: $\ell(wx_i, y_i) = |wx_i - y_i|$

$$\hat{L}(w) = |w - 1| + |2w - 3|$$

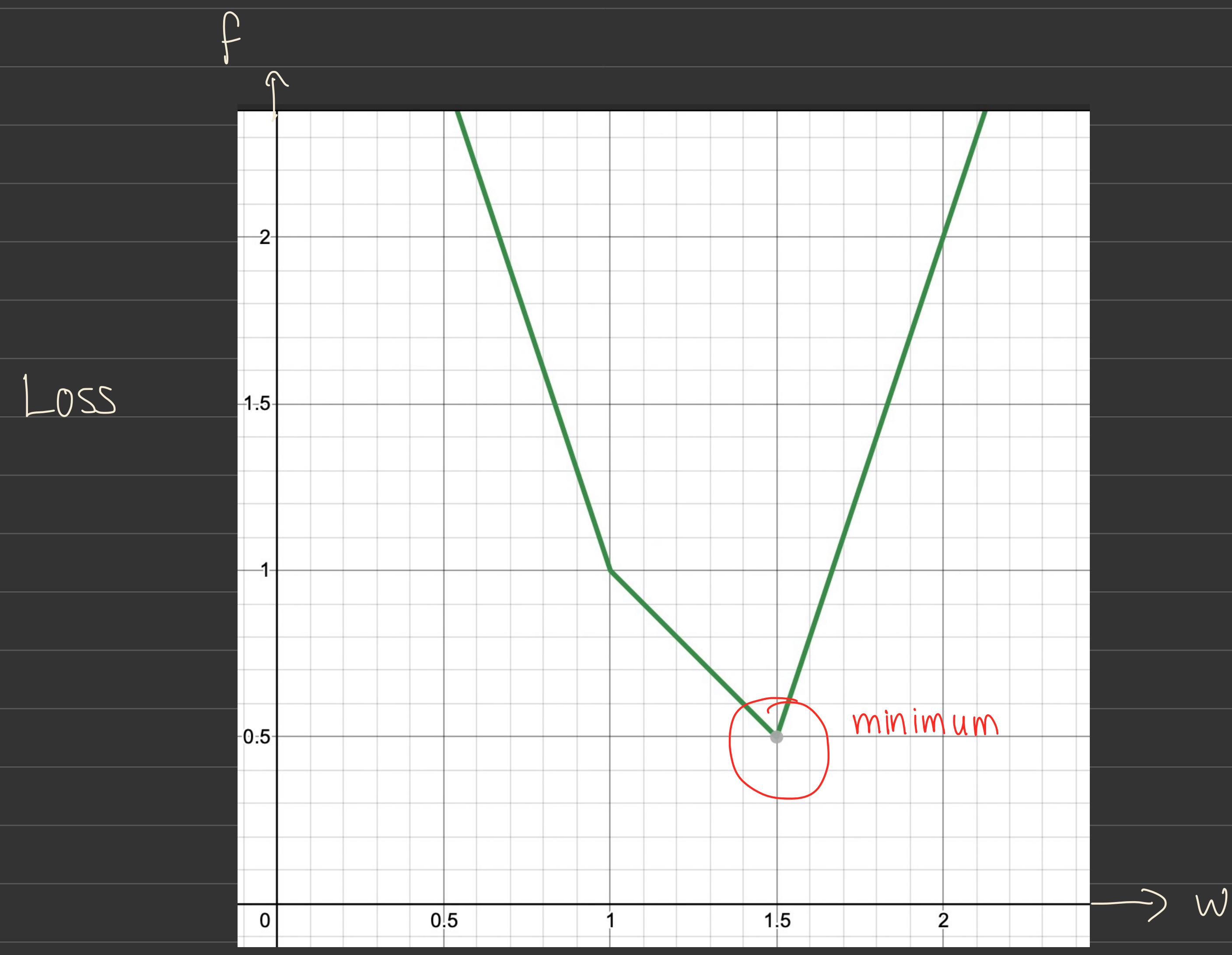
sign change at $w = 1$, $w = \frac{3}{2}$

For $w \le 1$: $f(w) = (1 - w) + (3 - 2w) = -3w + 4$

For $1 < w \le 3/2$ $f(w) = (w - 1) + (3 - 2w) = -w + 2$

For $w > 3/2$: $f(w) = (w - 1) + (2w - 3) = 3w - 4$

Piecewise linear & can plot it



$$\Rightarrow \boxed{w = 3/2}$$

(4) (Two variable quadratic regression) Set $X = \{(3,0),(0,2),(1,1)\}$ $Y = \{6,2,5\}$ Setup the quadratic regression problem (i) by minimizing (EL) directly (i.e. take derivatives with respect to w_1 and w_2 and solve. (ii) by setting up the matrix equation (10) and solving it.

$$X = \begin{array}{cc} \begin{bmatrix} 3 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \\ x_1 \quad x_2 \end{array} \qquad Y = \begin{bmatrix} 6 \\ 2 \\ 5 \end{bmatrix}$$

want to find $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

2 features now

(1) Taking derivatives & solving

$$\hat{L}(w) = \frac{1}{3}\left[\frac{(3w_1-6)^2}{2} + \frac{(2w_2-2)^2}{2} + \frac{(w_1+w_2-5)^2}{2}\right)$$

$$\frac{\partial L(w)}{\partial w_1} = \frac{1}{3}\left[3(3w_1-6) + (w_1+w_2-5)\right]$$

$$= \frac{1}{3}(10w_1 + w_2 - 23)$$

$$\frac{\partial L(w)}{\partial w_2} = \frac{1}{3}\left[2(2w_2-2) + (w_1+w_2-5)\right]$$

$$= \frac{1}{3}(w_1 + 5w_2 - 9)$$

Finding minimum by solving linear equations

$$10w_1 + w_2 - 23 = 0$$
$$w_1 + 5w_2 - 9 = 0$$

Substitution : see $w_1 = 9 - 5w_2$

$$10(9 - 5w_2) + w_2 - 23 = 0$$
$$\Rightarrow 90 - 50w_2 + w_2 - 23 = 0$$
$$\Rightarrow 67 - 49w_2 = 0$$
$$\Rightarrow \boxed{w_2 = 67/49}$$

$$w_1 = 9 - 5(67/49)$$
$$\Rightarrow \boxed{w_1 = 106/49}$$

② Solving the matrix equations we derived

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

$$= \left(\begin{bmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 2 \\ 5 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 1 \\ 1 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 23 \\ 9 \end{bmatrix}$$

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

$$A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$= \begin{bmatrix} 5/49 & -1/49 \\ -1/49 & 10/49 \end{bmatrix} \begin{bmatrix} 23 \\ 9 \end{bmatrix}$$

$$= \begin{bmatrix} 106/49 \\ 67/49 \end{bmatrix}$$