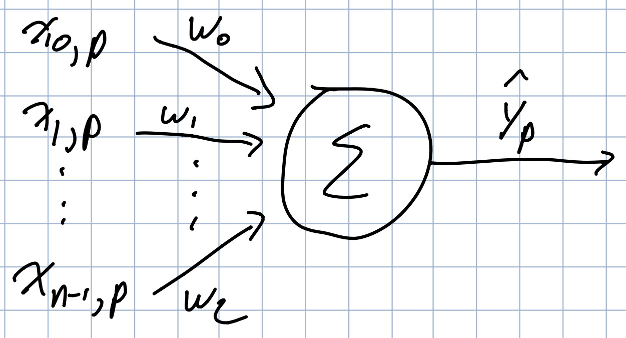

$$\hat{y}_p = \sum_{i=0}^{n-1} x_{i,p} w_i$$

$$p \in [0 \cdots m-1]$$

Training set: X_0, y_0
$\vdots$
X_{m-1}, y_{m-1}

E was formulated a little differently in the code than what we had been doing on the board:

$$E = \frac{1}{m}\sum_{p=0}^{m-1} E_p = \frac{1}{m}\sum_{p=0}^{m-1} (\hat{y}_p - y_p)^2$$

$$\frac{\partial E}{\partial w_i} = \frac{1}{m}\sum_p \frac{\partial}{\partial w_i}(\hat{y}_p - y_p)^2$$

$$= \frac{1}{m}\sum_p 2(\hat{y}_p - y_p)\frac{\partial}{\partial w_i}\hat{y}_p$$

$$= \frac{2}{m}\sum_p (\hat{y}_p - y_p)\, x_{i,p}$$

$$= \frac{2}{m}\sum_p \text{error}_p \cdot x_{i,p}$$

In the code, gradients is a vector:

$$\text{gradients} = \frac{\partial E}{\partial W} = \frac{2}{m}\begin{bmatrix} \sum_p e_p \cdot x_{0,p} \\ \sum_p e_p \cdot x_{1,p} \\ \vdots \\ \sum_p e_p \cdot x_{n-1,p} \end{bmatrix}$$

Part 1:

HW 1: add a bias term to our model:

$$\hat{y}_p = \left(\sum x_{i,p} \cdot w_i\right) + b$$

a) add a 'b' variable

b) make any changes to the forward model that are necessary

c) make any changes to 'gradients' that are necessary

d) compute a gradient for b

e) add a new training operation for b

f) run this training op in your learning loop.

Part 2: add regularization terms

$$E = \frac{1}{m}\sum_{p=0}^{m-1} E_p + \gamma\left(\sum_{i=0}^{n-1} w_i^2 + b^2\right)$$

a) modify the gradients computation

b) modify the gradients_bias computation

c) Select $\gamma = 2$