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

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

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

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

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

$$= \frac{2}{m} \sum_{p=0}^{m-1} \text{error}_p \cdot x_{i,p}$$

In the code, gradients is a vector:

$$\text{gradients} = \frac{dE}{dw} = \frac{2}{m} \begin{bmatrix} \sum_{p=0}^{m-1} \text{error}_p \cdot x_{0,p} \\
\vdots \\
\sum_{p=0}^{m-1} \text{error}_p \cdot x_{n-1,p} \end{bmatrix}$$
Part 1: add a bias term to our model:

\[ \hat{y}_p = (\sum x_{ip} w_i) + b \]

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

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

a) modify the gradients computation
b) modify the gradients-bias computation
e) Select \( \gamma = 2 \)