Objective function for linear regression:

$$C(\theta; x, y) = \sum_i (y^{(i)} - (\theta^T x^{(i)})^2.$$

Have the students work on obtaining the gradient (first for the 2D θ and then for a general θ for that before continuing. Perhaps try walking around the room to see how they're doing.

(Note that the y and x are fixed.)

For $dim(\theta) = 2$:

$$C(\theta; x, y) = \sum_i (y^{(i)} - (\theta_0 + \theta_1 x^{(i)})^2$$

Derivative (if we set $x_0 = 1$ – explain this, and probably use the case when $dim(\theta) = 2$ in the explanation), using the chain rule

$$\frac{\partial}{\partial \theta_i} C(\theta; x, y) = -2 \sum (y^{(i)} - \theta^T x^{(i)}) x^{(i)}$$

.

Stack those to get ∇C.

(I suspect many people will try to expand the square, which can work, although it's harder.)

Now work on vectorizing this code.

Now move on to modifying

`http://www.cs.toronto.edu/~guerzhoy/321/lec/W02/python/grad_descent2.py` in order to perform linear regression.

(Basically, the issue is that the parameters x and y are fixed.)

Try a smallish learning rate to see that it produces problems. Then try the small learning rate to see that it works.