Gradient Checking

We have a function $J(b, \omega)$ we want to minimize.

To do this, we need a function

$$dJ(b, \omega) = \begin{bmatrix} \frac{\partial J}{\partial b} \\ \frac{\partial J}{\partial \omega_1} \\ \vdots \\ \frac{\partial J}{\partial \omega_p} \end{bmatrix}$$

$dJ(b, \omega_1, \ldots, \omega_p)$

How can we check to be sure our implementation isn't buggy?

Compute $\dfrac{J(b+\varepsilon, \omega_1, \ldots, \omega_p) - J(b-\varepsilon, \omega_1, \ldots, \omega_p)}{2\varepsilon}$

If dJ is implemented correctly, the above should be approximately equal to

$$dJ(b, \omega_1, \ldots, \omega_p).$$