Explanatory Notes for 6.390

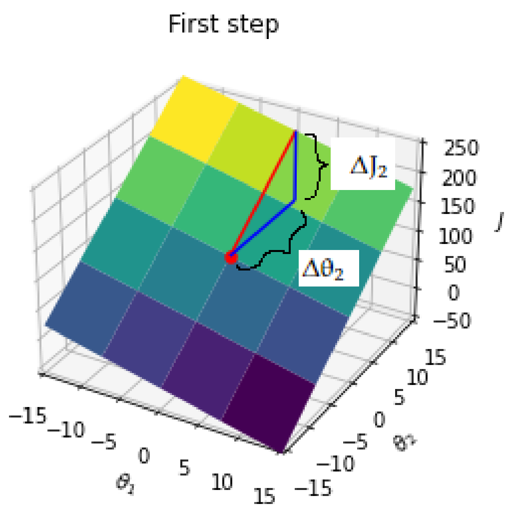
Shaunticlair Ruiz (Current TA)

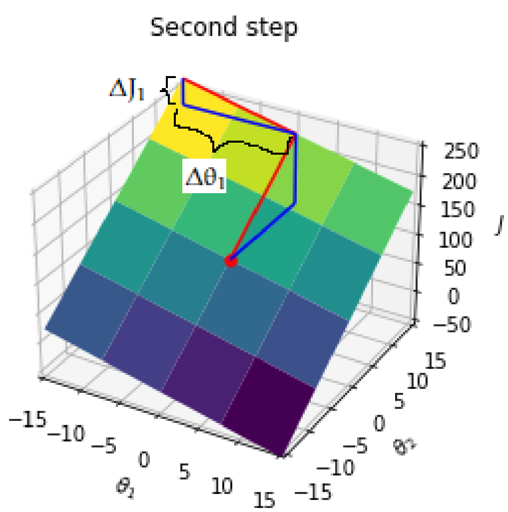
Fall 2022

The Optimal Direction: The Gradient

How do we get the optimal direction?

The **total** change in J is gotten by just **adding** the change in each direction (thank you planes!):





You can add up the results of our two steps: ΔJ_2 and ΔJ_1.

$$\Delta J \approx \Delta J_1 + \Delta J_2 \tag{1}$$

Let's convert that using derivatives:

$$\Delta J \approx \Delta\theta_1 \frac{\partial J}{\partial \theta_1} + \Delta\theta_2 \frac{\partial J}{\partial \theta_2} \tag{2}$$

Now we've got a useful equation: the total change. As a bonus we can see a clear **pattern**

(i^{th} θ matches i^{th} derivative).

So, **condense** this pattern, like we did for our linear model: using a **dot product**.

$$\Delta J \approx \begin{bmatrix} \Delta\theta_1 \\ \Delta\theta_2 \end{bmatrix} \cdot \begin{bmatrix} \partial J/\partial\theta_1 \\ \partial J/\partial\theta_2 \end{bmatrix} = \Delta\theta \cdot \nabla_\theta J \tag{3}$$

The **gradient** shows up! Interesting. But what does that **mean**?

Well, we want to **maximize** (or minimize!) our ΔJ. How do we maximize a **dot product**?

By making sure the directions are **the same**! So, we can confirm that the **gradient** gives us the **best** direction.

So, all we have to do is to **flip** the sign to **minimize** ΔJ.

And so, gradient descent is already complete!

Concept 1

The **gradient** ∇J is the **direction of greatest increase** for J.

That means means the opposite direction $-\nabla J$ is the **direction of greatest decrease** in J.

This is the single **most important concept** in this entire chapter!