## 13.2.3 Gradient descent in multiple dimensions

MO2.12

Now let's consider a problem with two dimensions, in other words, $J\left(a_0, a_1\right)$ is a function of two states, $a_0$ and $a_1$. We can use a Taylor series to approximate change in $J$ for a change in $\underline{a}=\left(a_0, a_1\right)$. Let's define the change in $\underline{a}$ as $\Delta\underline{a}=\left(\Delta a_0, \Delta a_1\right)$. The Taylor series for $J\left(a_0+\Delta a_0, a_1+\Delta a_1\right)$ about $J\left(a_0, a_1\right)$ is,

$$J\left(a_0+\Delta a_0, a_1+\Delta a_1\right) \approx J\left(a_0, a_1\right)+\Delta a_0 \left.\frac{\partial J}{\partial a_0}\right|_{\left(a_0, a_1\right)}+\Delta a_1 \left.\frac{\partial J}{\partial a_1}\right|_{\left(a_0, a_1\right)} \tag{13.14}$$

We can condense this Taylor series approximation using the gradient of $J$ which is defined as,

$$\nabla J\left(a_0, a_1\right)=\left(\left.\frac{\partial J}{\partial a_0}\right|_{\left(a_0, a_1\right)},\left.\frac{\partial J}{\partial a_1}\right|_{\left(a_0, a_1\right)}\right). \tag{13.15}$$

Thus, we can write Equation (13.14) as,

$$\Delta J \equiv J\left(\underline{a}+\Delta \underline{a}\right)-J\left(\underline{a}\right) \approx \Delta \underline{a} \cdot \nabla J\left(\underline{a}\right) \tag{13.16}$$

The notation $\underline{u} \cdot \underline{v}$ is for the dot product of $\underline{u}$ and $\underline{v}$.

This equation shows a few interesting facts about gradients:

- When the gradient is zero, i.e., all the partial derivatives of $J$ are zero, then the function values don't change to first order, so we are at an extremum.
- Also, note that when $\left(\Delta a_0\right)^2+\left(\Delta a_1\right)^2=1$ (the length squared of the $\Delta\underline{a}$ vector is one), then $\frac{\partial J}{\partial a_0} \Delta a_0+\frac{\partial J}{\partial a_1} \Delta a_1$ is the derivative of $J$ in the direction of $\Delta\underline{a}$.
- We can express the dot product as

$$\Delta \underline{a} \cdot \nabla J=\|\Delta \underline{a}\|\;\|\nabla J\|\;\cos \theta, \tag{13.17}$$

  where $\theta$ is the angle between the two vectors, and $\|\underline{v}\|=\sqrt{v_0^2+v_1^2}$ is the length (or norm) of a vector $\underline{v}$. To make the function increment $\Delta J$ largest for a $\Delta\underline{a}$ of a fixed length, we should pick

## Discussions

Posting in discussic

All posts sorted by recent activity

< Previous

Next >

## edX

- About
- Affiliates
- edX for Business
- Open edX
- Careers
- News

## Legal

- Terms of Service & Honor Code
- Privacy Policy
- Accessibility Policy
- Trademark Policy
- Sitemap
- Cookie Policy
- Your Privacy Choices

## Connect

- Idea Hub
- Contact Us
- Help Center
- Security
- Media Kit

GET IT ON Google play

Download on the App Store



**Video on derivation of multi-dimensional gradient descent**