COMS10003 Lecture 17.

Julian Gough 2015-2-23

Preface

These are outline notes for lecture 17.

Introduction

This is a lecture about partial differentiation and the gradient vector.

Partial differentiation

So, when we differentiate a function, we are working out its rate of change:

$$\frac{df}{dt} = \lim_{h \to 0} \frac{f(t+h) - f(t)}{h} \tag{1}$$

That is, it tells us how much $f(t)$ is increasing as t increases, at exactly the point t. Now, functions can depend on more than one variable, so we could have $f(x, y)$ for example. Now, at it's simplest, the partial derivative calculates how the function changes as one of the variables changes.

$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h} \tag{2}$$

and

$$\frac{\partial f}{\partial y} = \lim_{h \to 0} \frac{f(x, y+h) - f(x, y)}{h} \tag{3}$$

Note the partial sign, ∂, instead of the d, to denote partial differentiation. Make sure you write these two characters differently!

In practice, this means that when taking the partial derivative with respect to x you act as if y is a constant and visa versa. Hence, if

$$f(x, y) = xy^2 + x^2 + y \tag{4}$$

then

$$\frac{\partial f}{\partial x} = y^2 + 2x \tag{5}$$

and

$$\frac{\partial f}{\partial y} = 2xy + 1 \tag{6}$$

The gradient vector and the directional derivative

Often x and y aren't the only directions you are interested in and what you want is the *directional derivative*, the derivative in some other direction. The most convenient way to describe a direction mathematically is with a unit vector

$$\mathbf{n} = (n_1, n_2) \tag{7}$$

with $\sqrt{n_1^2 + n_2^2} = 1$. Here, I am using the bold notation for vectors, this is the most common notation although in school mathematics an arrow over the letter is common. One potentially confusing thing is that in hand-written mathematics the bold is indicated with an underline, so $\underline{n}$ in hand writing is the same as $\mathbf{n}$ in print, just like 'a' in print is the same as 'ɑ' in handwriting. Anyway, the idea is that a vector has direction and length, but a unit vector is always one long, so it is used to define direction.

Now, the directional derivative, written $\nabla_{\mathbf{n}}$ is a measure of how much f changes in the direction $\mathbf{n}$. We could write $\mathbf{x} = (x, y)$ and

$$\nabla_{\mathbf{n}} f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{n}) - f(\mathbf{x})}{h} \tag{8}$$

or, equivalently

$$\nabla_{\mathbf{n}} f(x, y) = \lim_{h \to 0} \frac{f(x + hn_1, y + hn_2) - f(x, y)}{h} \tag{9}$$

Luckily there is an easy formula for this

$$\nabla_{\mathbf{n}} f(\mathbf{x}) = \nabla f \cdot \mathbf{n} \tag{10}$$

where ∇f is the *gradient* of $f(x, y)$

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \tag{11}$$

Thus, the gradient of $f(x, y)$ is a vector made out of the partial derivatives. It is sometimes written as grad$f$ so

$$\text{grad} f = \nabla f \tag{12}$$

The formula for the directional derivative says it is the dot product of the gradient with the direction vector. To remind you, if $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ are two vectors then their dot product is

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 = |\mathbf{a}||\mathbf{b}| \cos\theta \tag{13}$$

where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. If $\mathbf{n}$ is a unit vector then $\mathbf{a} \cdot \mathbf{n}$ is a *projection* of $\mathbf{a}$ into the direction $\mathbf{n}$; it shows how much of $\mathbf{a}$ lies in the direction $\mathbf{n}$. In this way we can think of the gradient desribing how f is changing and the directional derivative as saying how much of that change is in the direction.

Lets do a quick example, consider

$$f(x, y) = x^2 + y^2 \tag{14}$$

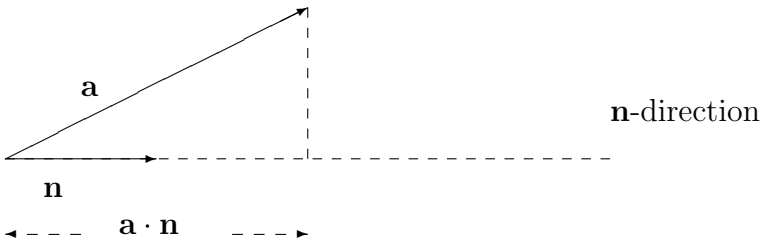

Figure 1: The dot-product: $\mathbf{a} \cdot \mathbf{n}$ gives the length of the perpendicular projection of $\mathbf{a}$ onto the direction indicated by the unit vector $\mathbf{n}$.

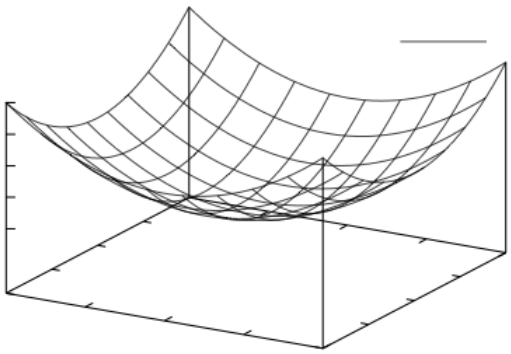

Figure 2: The function $x^2 + y^2$ is shaped like a bowl.

and say you want the derivative in the direction

$$\mathbf{n} = \frac{1}{\sqrt{2}}(1, -1) \tag{15}$$

at the point $(2, 2)$. Well

$$\nabla f = (2x, 2y) \tag{16}$$

so at $(2, 2)$ we have

$$\nabla f|_{(2,2)} = (4, 4) \tag{17}$$

and hence the directional derivative is

$$(1, -1) \cdot (4, 4) = 4 - 4 = 0 \tag{18}$$

This makes sense because if you think of the symmetry of the function it doesn't change if you go around.

The direction of change

Now, lets look at the directional derivative again

$$\nabla_{\mathbf{n}} f(x, y) = \nabla f \cdot \mathbf{n} = |\nabla f| \cos\theta \tag{19}$$

where, again, θ is the angle between ∇f and $\mathbf{n}$. Clearly this is biggest when $\theta = 0$, that is, when $\mathbf{n}$ is in the same direction as ∇f. Thus, ∇f points in the direction $f(x, y)$ is changing the fastest, the direction with the largest directional derivative. In the bowl example above, for example, $\nabla f = (2x, 2y)$, that is, it always points straight outwards, straight up the side of the bowl.

Without going into details here, gradients are important in some areas of computer science because of *gradient descent*. This is an approach to minimizing a function of lots of variables. In machine learning, for example, this function might be an error function, expressing how an estimate differs from the thing it estimates. In gradient descent the gradient vector is used to decide how to reduce the error by, as it were, heading straight down the gradient hill. In fact, the story is a bit more complicated, because of the hazard presented by thin valleys in the error landscape, a slightly different method is used, called *conjugate gradient*.

Other differential operators

The gradient is one of three common differential operators. Another common operator is the *divergence*; the gradient acts on scalar functions to give a vector, the divergence is the other way around, it acts on vectors to give a scalar function. Say $\mathbf{f}$ is a vector function, in two-dimensions it might be

$$\mathbf{f}(x, y) = (f_1(x, y), f_2(x, y)) \tag{20}$$

The divergence of $\mathbf{f}$ is then

$$\text{div}\mathbf{f} = \nabla \cdot \mathbf{f} = \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} \tag{21}$$

Here is a rough interpretation. Imagine the vector field is the velocity of a gas, in some places the vectors are getting shorter because it's somewhere the gas is slowing down, in others they are getting longer because the gas is speeding up. In some places the gas is spreading out, so the directions are changing. The divergence at a point will measure whether the density of the gas at the point is going up or down. It measures whether the velocity vector field is changing in such a way as would cause the gas to accumulate or its opposite.