DATA605

Week12+13 Lecture

Outline

1. Limits
    a. Intuitive notion
    b. Visualizations
    c. formalization
2. Continuity
3. Derivatives
    a. Intuitive notion
    b. Visualization
    c. Compute via limit
    d. Show common differential rules
4. Integrals
    a. Intuitive notion
    b. Visualization
    c. Show common integration rules
    d. Compute
5. Fundamental Theorem of Calculus
6. Multivariate Calculus and Partial Derivatives (2 slides)
7. Review homework
8. Bias and Variance Wrap Up

Limits: Intuition

Intuitive notion: When we write

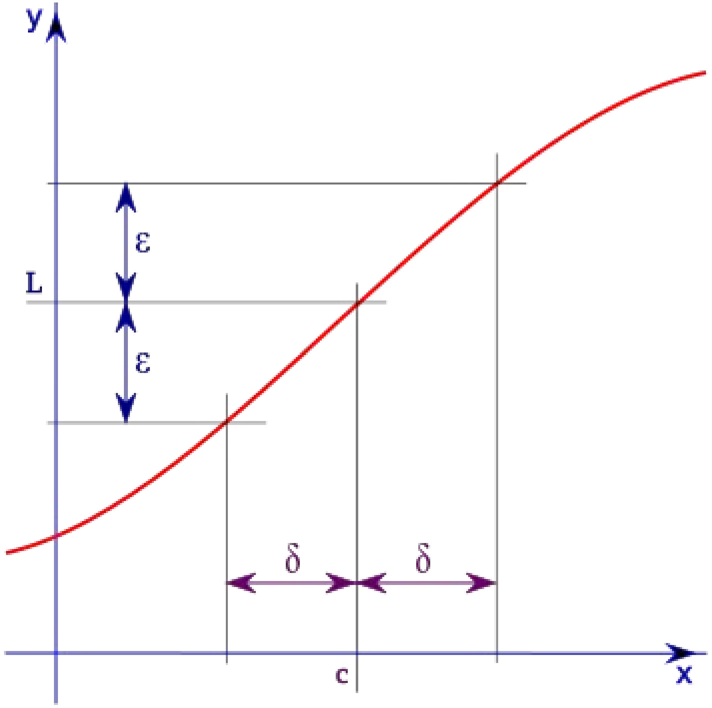
$\lim_{t \to \tau} v(t) = V$

what we are saying is that the function $v(t)$ gets arbitrarily close to V as t gets arbitrarily close to τ . We need to formalize this notion, but for now, we will accept the intuition that we are working on a small neighborhood around the point of interest and we examine how the function of interest converges (or diverges) around that point of interest.

Limits: Visualization

- http://webspace.ship.edu/msrenault/GeoGebraCalculus/limit_intuitive.html
- http://webspace.ship.edu/msrenault/GeoGebraCalculus/limit_intuitive_one_side.html

Limits Formalization



From https://en.wikipedia.org/wiki/Limit_(mathematics)

To formalize these notions, we need to figure out how we can represent 'arbitrarily close' from our intuitive definition. For example, "$f(x)$ becomes arbitrarily close to L" means that $f(x)$ eventually lies in the interval $(L - \varepsilon, L + \varepsilon)$, which can also be written using the absolute value sign as $|f(x) - L| < \varepsilon$.[2]

The phrase "as x approaches c" then indicates that we refer to values of x whose distance from c is less than some positive number δ (the lower case Greek letter *delta*)—that is, values of x within either $(c - \delta, c)$ or $(c, c + \delta)$, which can be expressed with $0 < |x - c| < \delta$.

The first inequality means that the distance between x and c is greater than 0 and that $x \neq c$, while the second indicates that x is within distance δ of c.

Therefore, our formal definition is:

$\lim_{x \to c} f(x) = L$ means that given any real number $\varepsilon > 0$, there exists another real number $\delta > 0$ so that if $0 < |x - c| < \delta$, then $|f(x) - L| < \varepsilon$

Limit Rules

Two rules of limits (not in the lecture notes):

Sum Law: $\lim_{x \to c} (f(x) + g(x)) = \lim_{x \to c} f(x) + \lim_{x \to c} g(x)$

Product Law: $\lim_{x \to c} (f(x)g(x)) = (\lim_{x \to c} f(x))(\lim_{x \to c} g(x))$

Continuity at a Point (Not in Lecture Notes)

- http://webspace.ship.edu/msrenault/GeoGebraCalculus/continuity_at_a_point.html

The function f is continuous at the x-value c if

1. $\lim_{x \to c} f(x) = L$ where L is a real number (so L can't be ∞ or $-\infty$), and
2. $f(c) = L$.

Informally, we say the function is continuous at a point if the limit exists, the function value exists, and they are equal to each other.

Derivatives: Intuition

Calculus is all about change. We would like to know how a function changes its output value as we change its input value. The derivative is a way of doing this, a "measures of sensitivity to change of the function (output) value with respect to a change in its argument (input value)"

From: https://en.wikipedia.org/wiki/Derivative

Derivatives: Visualization

- http://webspace.ship.edu/msrenault/GeoGebraCalculus/derivative_at_a_point.html
- http://webspace.ship.edu/msrenault/GeoGebraCalculus/derivative_as_a_function.html

Derivatives: Formalization

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

The derivative of a function therefore is simply a function of x that outputs the slope of the tangent line at a given point of that function. Note that f'(x) is spoken as 'f prime of x'.

First image: Derivative at a point x = a is the formal definition of the derivative at a point

Second image: Derivative as a function of all input values of x

Derivative: Rules

- http://webspace.ship.edu/msrenault/GeoGebraCalculus/derivative_elementary_functions.html

Common Function	Form	Differential Form
Constant	$f(x) = c$	$f'(x) = 0$
Identity	$f(x) = x$	$f'(x) = 1$
Square	$f(x) = x^2$	$f'(x) = 2x$
Square root	$f(x) = \sqrt(x)$	$f'(x) = \frac{1}{2}x^{-\frac{1}{2}}$
General form of the above	$f(x) = x^n$	$f'(x) = nx^{n-1}$
Exponential	$f(x) = e^x$	$f'(x) = e^x$
	$f(x) = a^x$	$a^x(ln(a))$
Logarithm	$f(x) = ln(x)$	$f'(x) = \frac{1}{x}$
Trigonometry	$f(x) = \sin(x)$	$f'(x) = \cos(x)$
	$f(x) = \cos(x)$	$f'(x) = -\sin(x)$

Rule	Function	Differential Form
Multiplication by constant	$cf(x)$	$cf'(x)$
Sum of Functions	$f(x) + g(x)$	$f'(x) + g'(x)$
Product of Functions	$f(x)g(x)$	$f'(x)g(x) + g'(x)f(x)$
Reciprocal	$1/f(x)$	$-f'(x)/f^2(x)$
Quotient	$f(x)/g(x)$	$(f'(x)g(x) - g'(x)f(x))/g^2(x)$
Chain rule	$f(g(x))$	$f'(g(x))g'(x)$

Derivative: Chain Rule

$$\frac{d}{dx} f(g(x)) = f'(g(x))\, g'(x)$$

In the previous image, the chain rule is mentioned in passing. It is important to know.

For a visualization of this, check out: http://webspace.ship.edu/msrenault/GeoGebraCalculus/derivative_intuitive_chain_rule.html. The link also has some worked out examples.

Derivative: Numerical

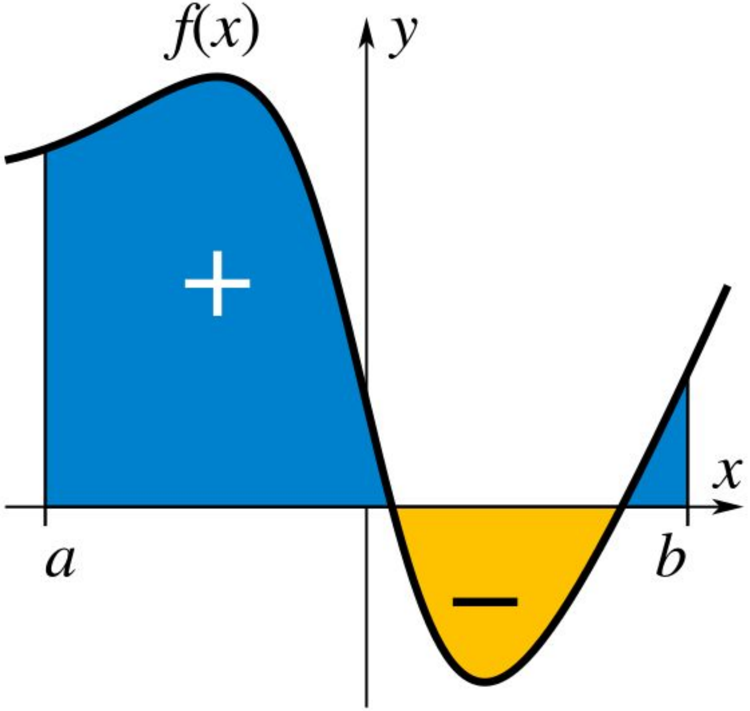
Why do we need numerical derivatives? Sometimes the analytical form is not known or can't be figured out since h = 0 in the denominator is undefined.

For the assignment this week, you need to find a derivative of a function numerically. Meaning, do not solve for the derivative in analytical' form. Your job is to write a function that takes in a value for x, and returns an approx value of the derivative of the given f(x) around that point.

Main idea: calculate derivative using the formalization, shrink h until the difference between your approximations of the derivative fall under some small threshold.

Another note: good programmers make sure to check their work on multiple inputs.

Integration: Intuition



We would also like to find the area under a function between two endpoints. Here is a diagram from wikipedia showing what we are after for a function f(x) between points a and b. This is what we call integration.

Notice: when we are above the x-axis, the area is considered positive; and considered negative when below the x-axis

Integration: Visualization

- http://webspace.ship.edu/msrenault/GeoGebraCalculus/integration_riemann_sum.html
- http://webspace.ship.edu/msrenault/GeoGebraCalculus/integration_area_function.html

Integration: Formalization

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\Delta x$$

Notice we need limits to function as our bedrock of both derivatives and integrals. The term inside the summation measures the area of those rectangles (Δx is the width, f(x) is the height) from the visualization. This area is an approximation of the area under the curve, but as n gets larger, that approximation gets better and better.

We say that "in the limit, the approximation becomes the exact area under the curve"

Integration: Rules

Before we go over rules and methods of integration, we should note that derivatives are a science, integrals are an art. We have pretty hard rules on how to calculate derivatives; integrals do have rules but many times we have to use ‘methods’ and knowing which method to use is sometimes difficult.

Rules of integration help you find the analytical form of the integral of a function.

Integration: Rules

Form	Integral Form
$\int a dx$	$ax + C$
$\int 2x dx$	$x^2 + C$
$\int x dx$	$x^2/2 + C$
$\int x^n dx$	$x^{n+1}/(n+1) + C$
$\int e^x dx$	$e^x + C$
$\int (1/x) dx$	$ln\|x\| + C$

Rule	Function	Integral Form
Multiply by Constant	$\int cf(x)dx$	$c\int f(x)dx$
Sum of Functions	$\int [f(x) + g(x)]dx$	$\int f(x)dx + \int g(x)dx$

Fundamental Theorem of Calculus

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x)$$

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

The first part says that integration and derivatives are in some sense complementary operations.

The second part says that to calculate an integral, we need F(x), the anti-derivative of f(x).

Integration by Parts

$$\int u\,dv = uv - \int v\,du.$$

When faced with some complicated expressions, sometimes a straightforward rule will not work. Integration by parts is a strategy for dealing with integrals of the form to the left. Choosing u and dv is important!

Some examples on wikipedia as well, which are pretty good despite it being wikipedia. Sometimes you will need to use this method more than once, since the integral on the right hand side still needs to be evaluated.

Integration by Parts

LIATE rule [edit]

A rule of thumb proposed by Herbert Kasube advises that whichever function comes first in the following list should be chosen as *u*:[2]

- **L** - Logarithmic Functions: $\ln(x),\ \log_b(x)$, etc.
- **I** - Inverse trigonometric functions: $\arctan(x),\ \operatorname{arcsec}(x)$, etc.
- **A** - Algebraic functions: $x^2,\ 3x^{50}$, etc.
- **T** - Trigonometric functions: $\sin(x),\ \tan(x)$, etc.
- **E** - Exponential functions: $e^x,\ 19^x$, etc.

The function which is to be *dv* is whichever comes last in the list: functions lower on the list have easier antiderivatives than the functions above ther

Strategy: In general, one tries to choose *u* and *dv* such that *du* is simpler than *u* and *dv* is easy to integrate. Sometimes the above doesn't work, so remember it is a 'rule of thumb', or a 'heuristic'.

Multivariate Calculus and Partial Derivatives

$$f(x_1, x_2) = a_1x_1 + a_2x_2 + a_{12}x_1x_2$$

$$\frac{\partial f}{\partial x_1} = a_1 + a_{12}x_2$$

$$\frac{\partial f}{\partial x_2} = a_2 + a_{12}x_1$$

We do not cover too much multivariate calculus but we need to know about partial derivatives to understand gradients and the algorithm known as Gradient Descent (which is super important for deep learning and many ML models). IF we have a function of multiple variables, how do we take a derivative?

When dealing with multiple variables, we need to look at the derivative of the function with respect to individual input variables, or ‘partial’ derivatives.

From the text, we have the example on the left, which is a function of two variables. To take the partial derivative of f WRT x_1, treat x_2 as a constant. Therefore the term a_2x_2 is essentially a constant, and a derivative of a constant is 0 (hence it drops out).

Gradients

$$\Delta f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

The Gradient of a function is simply a vector with n dimensions, where n is the number of input variables for f. The vector components are the partial derivatives of $f(x_1, x_2, \ldots x_n)$ WRT the input variables.

Bias and Variance Tradeoff

What is the problem? From wikipedia: “problem of simultaneously minimizing two sources of error that prevent supervised learning algorithms from **generalizing** beyond their training set”. This error has three components:

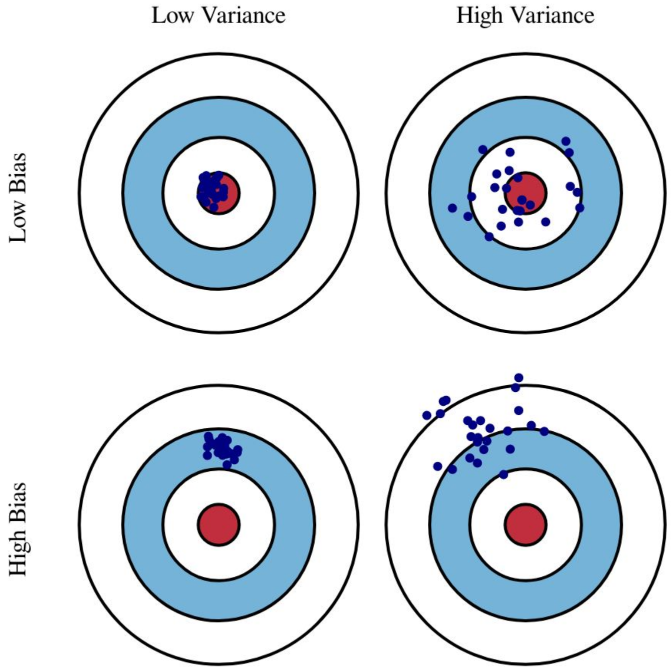
Irreducible Error: This is error that we cannot model effectively; this could be due to random noise in the samples.

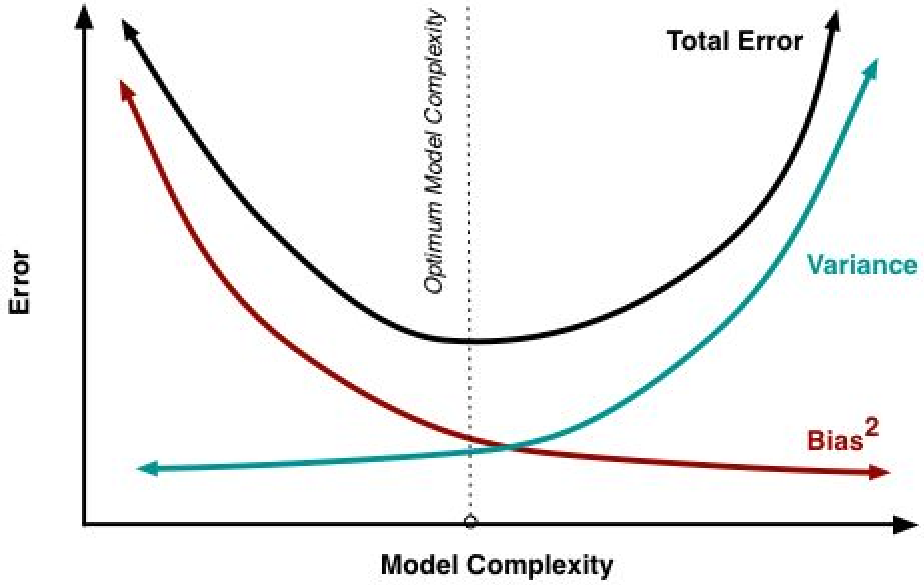
Bias: Error due to how far our model is from the real relationship between variables. This can be due to bad assumptions about the real relationship. Another way of saying it is the systematic departure of the model from the underlying true relationship

Variance: Error from sensitivity to small fluctuations in the training set.

Bias And Variance Tradeoff

Review http://scott.fortmann-roe.com/docs/BiasVariance.html





Assignment

A few questions from the assignment:

- There is variance due to the cross-validation, so not every run gives the same result
- It seems that most minimal errors happened from n = 2 to 4
  - Intuition: this makes sense, since this data comes from a physical system, and physical laws tend to not have variables to the 10th power
- The instructions should have said to go up to 10, instead of 8, which I think shows the U shape better
- Discussion board topic: why more variance in the model predictions?
  - https://stats.stackexchange.com/questions/144245/how-can-increasing-the-dimension-increase-the-variance-without-increasing-the-bi
  - Spatial Intuition: if you have the same amount of data points, but you are adding dimensions, you will naturally become more sensitive since data is more sparse in bigger input / sample space