Discussion 13

11 Dec 23

Announcements

- Homework 10 due on Wednesday
- OH today from 5 to 7pm at CCDS B64
- Reach out to Prof Pawel for a time conflict for the final

Review of gradients

The idea behind partial derivatives is to isolate each variable and then calculate the gradient.

Now we know how the function is varying along a single dimension. To get a sense of how the function is varying across all dimensions simultaneously, we collect all the partial derivatives of the function into a vector called the gradient.

Compute the gradient of the function $f(x) = x^2 + 3x + 2$.

Compute the gradient of the function $f(x) = x^2 + 3x + 2$.

To compute the gradient, we find it's derivative w.r.t. x.

$$f'(x) = 2x + 3$$

Since it only has 1 variable, this is the gradient.

Compute the gradient of the function $f(x, y) = xy^2 + 2x^2y - 4xy$.

Compute the gradient of the function $f(x,y) = xy^2 + 2x^2y - 4xy$.

Since it has 2 variables, we will compute the partial derivative of x and y and store it in a matrix.

Partial derivative w.r.t. x is given by:

$$f'(x,y) = y^2 + 4xy - 4y$$

Partial derivative w.r.t. y is given by:

$$f'(x,y) = 2xy + 2x^2 - 4x$$

Putting this in a matrix,

$$\nabla = \begin{bmatrix} y^2 + 4xy - 4y \\ 2xy + 2x^2 - 4x \end{bmatrix}$$

Let $f(x, y) = x^2 + y^2$. Find the gradient of f at the point $(1, 2)$.

Let $f(x, y) = x^2 + y^2$. Find the gradient of f at the point $(1, 2)$.

First, let's compute the gradient.

$$\nabla = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

Evaluating at point $(1, 2)$:

$$\nabla = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

Let $f(x, y, z) = x^2 + y^2 + z^2$. Find the partial derivatives of f with respect to x, y, and z.

Let $f(x, y, z) = x^2 + y^2 + z^2$. Find the partial derivatives of f with respect to x, y, and z.

Partial derivative w.r.t. x = $2x$

Partial derivative w.r.t. y = $2y$

Partial derivative w.r.t. z = $2z$

Let $f(x) = x^3 - 2x^2 + 3x$. Find the minimum of f using gradient descent.

Alpha = 0.01

X_0 = 1

Let $f(x) = x^3 - 2x^2 + 3x$. Find the minimum of f using gradient descent.

Finding the gradient,

$$f'(x) = 3x^2 - 4x$$

Using a learning rate $\alpha = 0.01$,

and the initial value of $x_0 = 1$,

Since $f'(x_0) = 2$,

$$x_1 = x_0 - \alpha(x_0) = 0.98$$

Now, $f'(x_1) = 1.96$, so,

$$x_2 = 0.9604$$

Now, $f'(x_2) = 1.92$,

$$x_3 = 0.94$$

Use gradient descent to find the minimum of the function $f(x) = x^2 - 4x + 3$.

Alpha = 0.01

X_0 = 3

Use gradient descent to find the minimum of the function $f(x) = x^2 - 4x + 3$.

Finding the gradient,

$$f'(x) = 2x - 4$$

Using a learning rate $\alpha = 0.01$,

and the initial value of $x_0 = 3$,

Since $f'(x_0) = 2$,

$$x_1 = x_0 - \alpha(x_0) = 2.94$$

Similarly, we have,

$$x_2 = 2.88$$

$$x_3 = 2.83$$

Use gradient descent to find the minimum of the function $f(x, y) = x^2 + y^2$.

X and y have initial values of 1

Learning rate = 0.01

2 iterations

Finding the gradient,

$$\nabla = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

Since this has 2 variables, we will simultaneously update the values of x and y. For this question, we will just perform 2 iterations.

We will use a learning rate of $\alpha = 0.01$ for this problem.

$$f'_x(x, y) = 2x$$

If $x_0 = 1$,

$$f'_{x_0}(x, y) = 2$$

Using the gradient descent algorithm,

$$x_1 = x_0 - \alpha f'_{x_0}(x, y)$$

$$x_1 = 1 - 0.01(2) = 0.98$$

Since the gradient descent algorithm requires a simultaneous update,

$$f'_y(x, y) = 2y$$

If $y_0 = 1$,

$$f'_{y_0}(x, y) = 2$$

Using the gradient descent algorithm,

$$y_1 = y_0 - \alpha f'_{y_0}(x, y)$$

$$y_1 = 1 - 0.01(2) = 0.98$$

Now, we have $x_1 = 0.98$ and $y_1 = 0.98$.

Note: It is just a coincidence that x and y will have the same values after every update for this function.

Iteration 2:

$$f'_x(x, y) = 2x$$

If $x_1 = 0.98$,

$$f'_{x_1}(x, y) = 1.96$$

Using the gradient descent algorithm,

$$x_2 = x_1 - \alpha f'_{x_1}(x, y)$$

$$x_2 = 0.98 - 0.01(1.96) = 0.96$$

Similarly, $y_2 = 0.96$.

Show that the gradient of the function $f(x, y) = x^2 + y^2$ is $\nabla f = (2x, 2y)$.

Show that the gradient of the function $f(x, y) = x^2 + y^2$ is $\nabla f = (2x, 2y)$.

This is a fairly straightforward question.

Partial derivative w.r.t. x:

$$f'_x(x, y) = 2x$$

Partial derivative w.r.t. y:

$$f'_y(x, y) = 2y$$

Hence,

$$f(x, y) = x^2 + y^2$$

Review of gradients

1. Given a function:

$$f(x, y) = x^5 + 3x^3y^2 + 3xy^4$$

Find it's partial derivatives.

2. Find the gradient of the above function.

3. Compute $x^T A x$ for the following matrix:

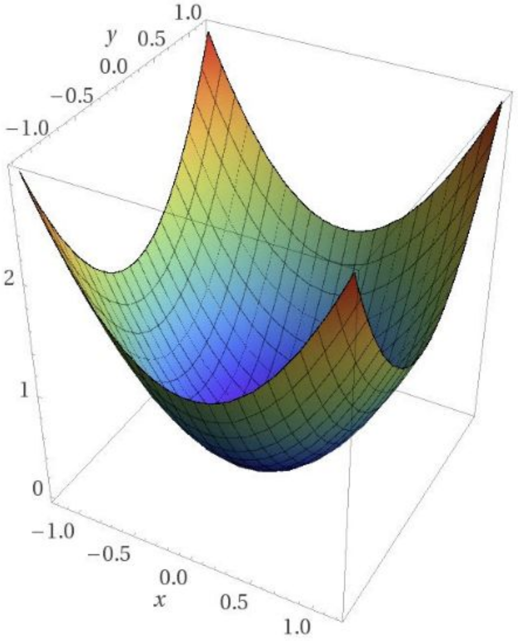
$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

The idea behind partial derivatives is to isolate each variable and then calculate the gradient.

Now we know how the function is varying along a single dimension. To get a sense of how the function is varying across all dimensions simultaneously, we collect all the partial derivatives of the function into a vector called the gradient.

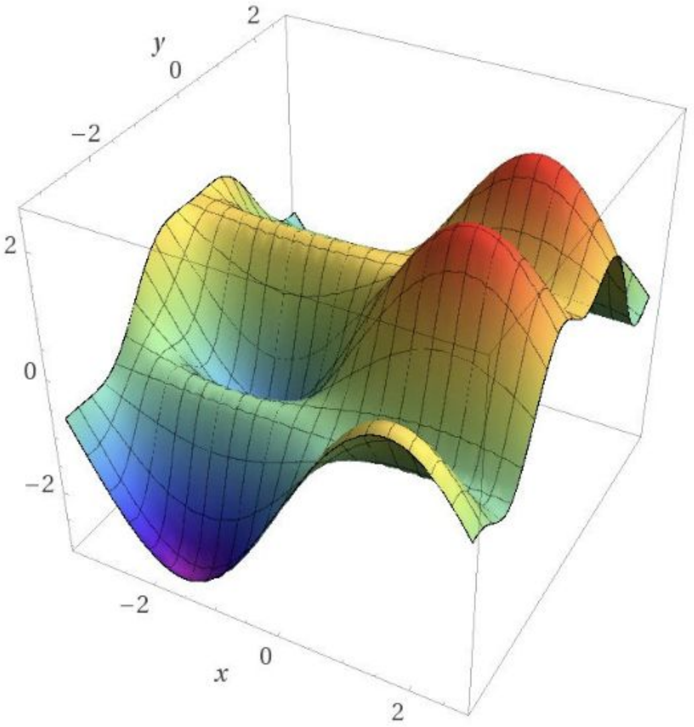
Gradient descent



1. Calculated in cases where we don't have a closed form solution
2. Can be compared to going down a mountain based on the level of steepness
3. The direction of steepness is the gradient
4. The step size to go downhill is comparable to the learning rate

When is a gradient 0?

What about a non-convex function?



Which of the following seems to be true for a non-convex function:

1. It has no minima
2. It has 1 local minimum
3. It has multiple minima

Overview for the finals

- Basic probability - Conditional probability, CDF, PDF, PMF, distributions
- Hypothesis testing - Random variables, LOTUS, CLT, Sampling, confidence intervals, one tailed and two tailed tests, error rate
- Parameter estimation - point estimators, bias, variance, MSE, Bessel's correction, MLE
- Bayes, Bayes table, Bayes with distributions, Poisson processes, probability of superiority, Bayesian testing
- Conjugate priors
- Markov chains, state transition diagrams, MCMC, Metropolis-Hasting
- Hidden Markov Models, Viterbi algorithm
- Gradient Descent

Thank you