MA5832 Data Mining and Machine Learning

Week 2

Hong-Bin Liu

James Cook University

06 May 2020

Admistration

- Assessment 1: 20%, Due date: Week 2 - Sunday, 17th May 2020, 11:59pm AEST.
- Future sessions will be held on Thursday, 6:00pm AEST.

Outline

1. Probability
2. Optimisation
3. Demo
4. Questions?

Notations

- $p(a)$: Probability distribution of random variable a
- $p(a, b)$: Joint Probability distribution of two random variables
- $p(a|b)$: Conditional Probability distribution

Product Rule and Bayes' Rule for Conditional Dependent Variables

$p(x, y) = p(x) \, p(y)$

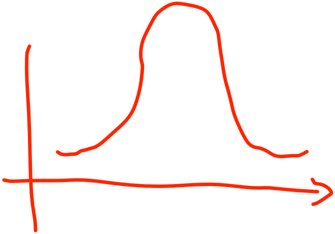
- Product rule: $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$

  $p(y)$

- Bayes' Rule: $\underbrace{p(\mathbf{x}|\mathbf{y})}_{\text{posterior}} = \dfrac{\overbrace{p(\mathbf{y}|\mathbf{x})}^{\text{likelihood}} \; \overbrace{p(\mathbf{x})}^{\text{prior}}}{\underbrace{p(\mathbf{y})}_{\text{evidence}}}$

Distributions

- Probability mass functions: Discrete probability distributions
- Probability density functions: Continuous probability distributions



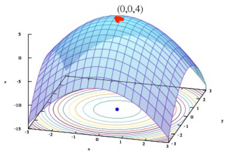
References

- "Mathematics for Machine Learning". Copyright 2020 by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. https://mml-book.com
- Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep learning. Cambridge, MA: MIT Press.Chapter 2: Linear Algebra (pp. 29-50). https://www.deeplearningbook.org/contents/prob.html

Outline

What is optimisation?

In the simplest case, an optimisation problem consists of maximising or minimising a real function by systematically choosing input values from within an allowed set and computing the value of the function.

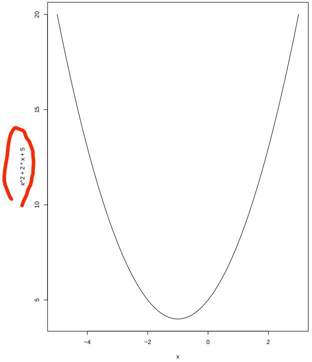


Graph of a given by $z = f(x, y) = -(x^2 + y^2) + 4$. The global maximum at $(x, y, z) = (0, 0, 4)$ is indicated by a blue dot

Figure: Taken from wikipedia.

Approaches

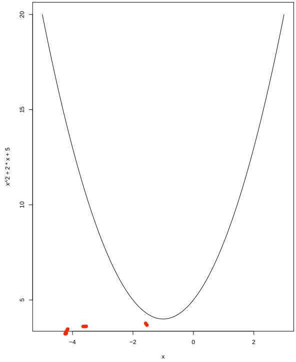
- Mathematical



$x^2 + 2x + 5$

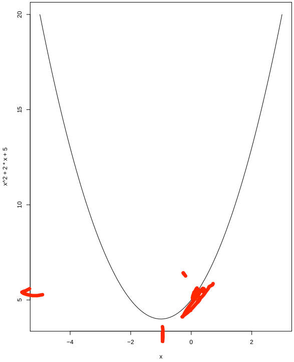
Approaches

- Mathematical
- Random search



Approaches

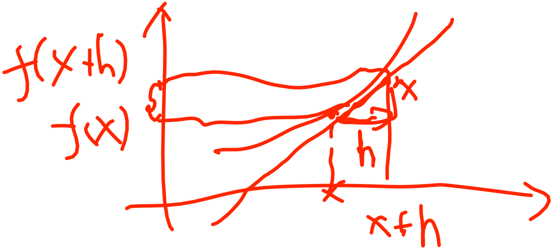
- Mathematical
- Random search
- Gradient-based methods



First Derivative

Definition 5.2 (Derivative). More formally, for $h > 0$ the derivative of f at x is defined as the limit

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



Derivatives of Common functions

- $(c)' = 0$
- $(x^a)' = ax^{a-1}$
- $(e^x)' = e^x$
- $(\sin x)' = \cos x$
- $(\cos x)' = -\sin x$

Differentiation Rules

x^{-1} x'

Product rule: $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$

Quotient rule: $\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$

Sum rule: $(f(x) + g(x))' = f'(x) + g'(x)$

Chain rule: $(g(f(x)))' = g'(f(x))f'(x)$

Second Derivatives

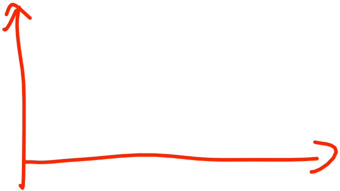
Second Derivatives is the derivative of derivative.

Example: $f(x) = x^3$

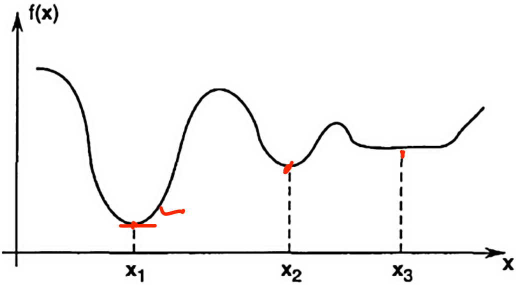
Its derivative is $f'(x) = 3x^2$ $\quad 3x^{3-1}$

The derivative of $3x^2$ is $6x$, so the second derivative of $f(x)$ is:

$$f''(x) = 6x$$

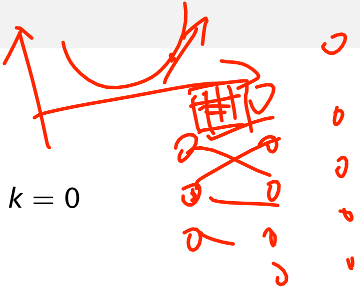


Global and local minima



Gradient Decent

Wx

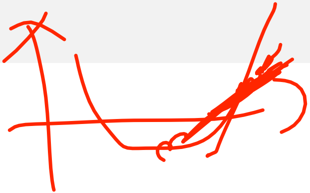


- Step 1. Given a starting point $x^{(k)}$, set $k = 0$
- Step 2. Find the gradient $\nabla f\left(x^{(k)}\right)$
- Step 3. Then find x^{k+1}

$$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f\left(x^{(k)}\right)$$

- Step 4. Set $k = k + 1$, repeat steps 2 to 4 a large number of times

Stochastic Gradient Descent (SGD)



- Step 1. Given a starting point $x^{(k)}$, set $k = 0$
- Step 2. Find the gradient $\nabla f\left(x^{(k)}\right)$ using subset
- Step 3. Then find x^{k+1}

$$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f\left(x^{(k)}\right)$$

- Step 4. Set $k = k + 1$, repeat steps 2 to 4 a large number of times

Newton's method

- Step 1. Given a starting point $x^{(k)}$, set $k = 0$
- Step 2. Find the gradient $\nabla f\left(x^{(k)}\right)$
- Step 3. Find the Hessian matrix $F\left(x^{(k)}\right)$
- Step 4. Then find x^{k+1} :

$$x^{(k+1)} = x^{(k)} - F\left(x^{(k)}\right)^{-1} \nabla f\left(x^{(k)}\right)$$

- Step 5. Set $k = k + 1$, repeat steps 2 through 5 a large number of times

JAMES COOK UNIVERSITY AUSTRALIA

Outline

Outline

1. Probability
2. Optimisation
3. Demo
4. Questions?

Questions?

Thank You.