INFO251 – Applied Machine Learning

Lab 5

Suraj R. Nair

Announcements

- **PS3 due tomorrow (Thursday, 2/20)**
  - Suraj Office Hours: Today (2/19) 10.30 – 12 noon, South Hall 107
  - Satej Office Hours: Tomorrow (2/20) 12.45 – 1.45, Room 6
- **PS4 will be released soon – update on submission format**
- **Quiz 1 on March 4 (next week's lab will be a review session)**
  - Suraj Extra Office Hours March 3 (time / location tbd)
- **Today: Gradient Descent**

Topics

- Gradient descent
  - Random initialization, learning rate, iterations, stopping conditions

Gradient Descent

1. Begin at a random point
2. Calculate the function value at the point and the gradient (partial derivatives)
3. Pick a new point, by moving in the opposite direction of the gradient. The size of the step is governed by the **learning rate**.
4. Repeat!

$$\mathbf{b} = \mathbf{a} - \gamma \nabla \mathbf{f}(\mathbf{a})$$

Review

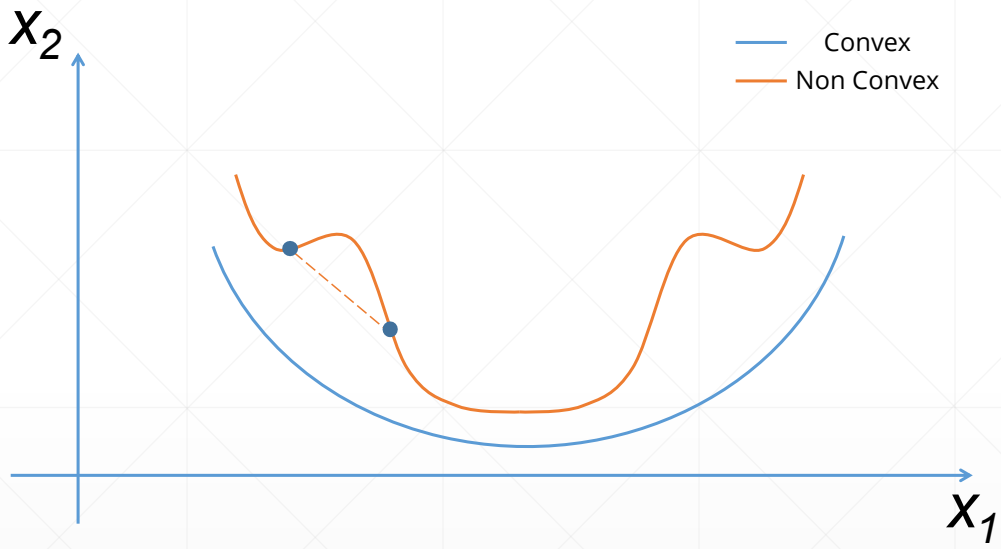
- You are trying to find the parameters for a multivariate linear regression using gradient descent. The algorithm is initialized at some random starting point. However, it is taking very long to converge (e.g > 10,000 iterations). What could be the reason(s)?
  - Step size is too small
  - Step size is too large
  - Data may not have been scaled
  - All of the above

Optimization

- Methods for optimization
  - Naïve grid search
  - Gradient descent
  - Linear programming, quadratic programming
  - Newton's method
  - …many, many more
- More on optimization**: EECS 127** and **EECS 227**

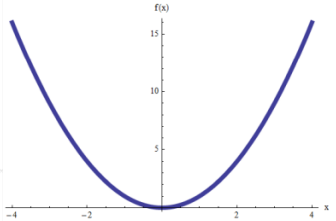
Convexity

- **Graphical interpretation**:
  - Line segment between any two points on the graph of the function does not lie below the graph
- **Convex function** (if twice differentiable):
  - **Single variable:** second derivative is always nonnegative
  - **Multivariable**: hessian matrix of second partial derivatives is positive semidefinite

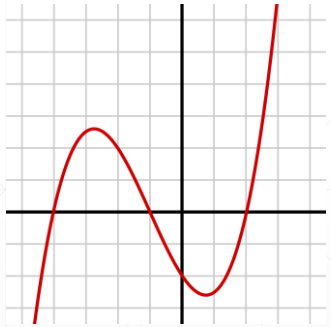


Convexity

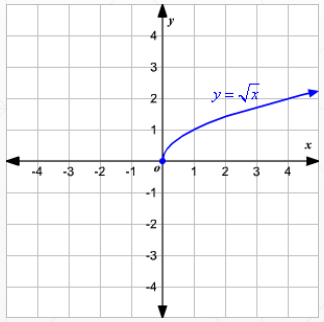
$f(x) = x^2$



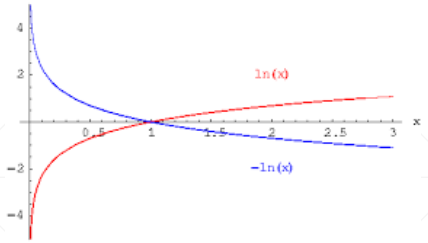
$f(x) = x^3$



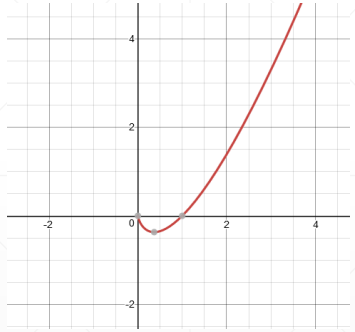
$f(x) = x^{1/2}$



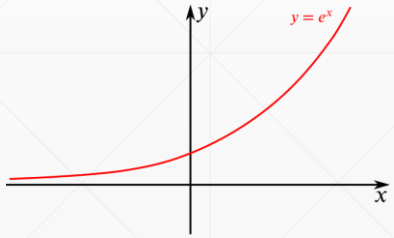
$f(x) = \ln(x)$



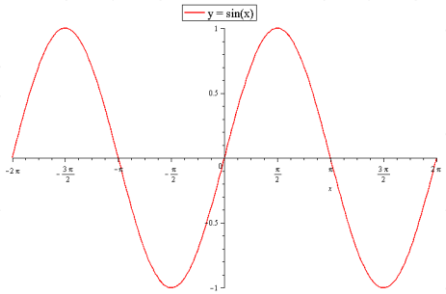
$f(x) = x\ln(x)$



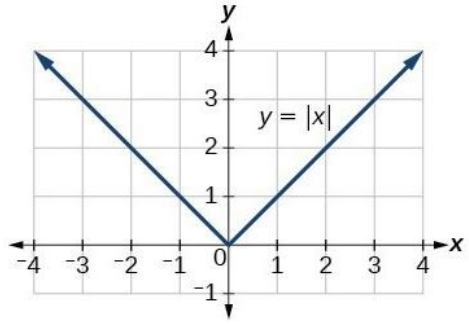
$f(x) = e^x$



$f(x) = \sin(x)$



$f(x) = |x|$



$f(x) = ax^2 - by^2$