Logistic Regression

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

- Learn the logistic regression classifier
  - Understand why the logistic regression is better suited than the linear regression for classification tasks
  - Understand the logistic function
  - Understand the logistic regression classifier
  - Understand the approximation approach for the open form solutions
- Learn the gradient descent algorithm
  - Know the tailor expansion
  - Understand the gradient descent/ascent algorithm
- Learn the different between the naïve Bayes and the logistic regression
  - Understand the similarity of the two classifiers
  - Understand the differences of the two classifiers
  - Understand the performance differences

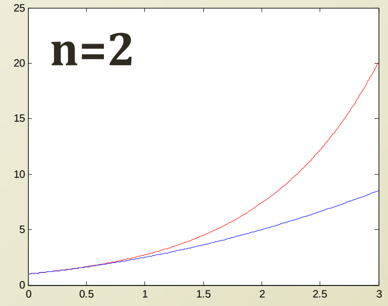
GRADIENT METHOD

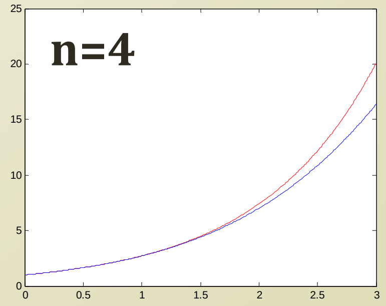
Taylor Expansion

- Taylor series is a representation of a function
  - as a infinite sum of terms calculated from the values of the function's derivatives at a fixed point.
  - $f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots$
    $= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$
  - a = a constant value
- Taylor series is possible when
  - Infinitely differentiable at a real or complex number of a
- Taylor expansion is a process of generating the Taylor series

$when\ a = 0,$

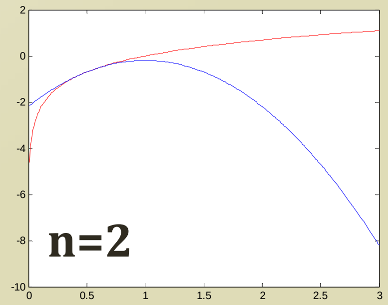
$$e^x = 1 + \frac{e^0}{1!}(x-0)^1 + \frac{e^0}{2!}(x-0)^2 + \cdots$$

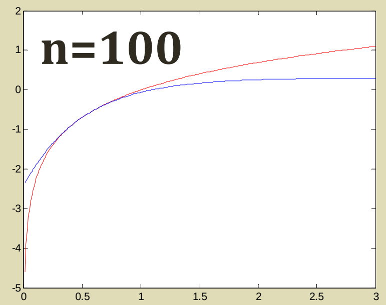




$when\ a = 0.5,$

$$logx = \log(0.5) + \frac{\frac{1}{0.5}}{1!}(x-0.5)^1 + \frac{\frac{1}{0.5^2}}{2!}(x-0.5)^2 + \cdots$$





Gradient Descent/Ascent

- Gradient descent/ascent method is
  - Given a differentiable function of *f(x)* and an initial parameter of x_1
  - Iteratively moving the parameter to the lower/higher value of *f(x)*
  - By taking the direction of the negative/positive gradient of *f(x)*
- Why this works?
  - $f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + O(||x-a||^2)$ Useful Big-Oh Notation
    - Assume $a=x_1$ and $x=x_1+h\mathbf{u}$, **u** is the unit direction vector for the partial deriv.
    - $f(x_1 + h\mathbf{u}) = f(x_1) + hf'(x_1)\mathbf{u} + h^2O(1)$
    - $f(x_1 + h\mathbf{u}) - f(x_1) \approx hf'(x_1)\mathbf{u}$ Always???
    - $\mathbf{u}^* = argmin_{\mathbf{u}}\{f(x_1 + h\mathbf{u}) - f(x_1)\} = argmin_{\mathbf{u}} hf'(x_1)\mathbf{u} = -\frac{f'(x_1)}{|f'(x_1)|}$
    - $\because f(x_1 + h\mathbf{u}) \le f(x_1), \vec{a}\cdot\vec{b} = |\vec{a}||\vec{b}|cos\alpha$ Gradient Descent
  - $x_{t+1} \leftarrow x_t + h\mathbf{u}^* = x_t - h\frac{f'(x_1)}{|f'(x_1)|}$
- Perfectly applicable to $\hat{\theta} = argmax_\theta \sum_{1\le i\le N} log(P(Y_i|X_i;\theta))$ Gradient Ascent
  - $f(\theta) = \sum_{1\le i\le N} log(P(Y_i|X_i;\theta))$
  - Setup an initial parameter of θ_1
  - Iteratively moving θ_t to the higher value of $f(\theta_t)$
  - By taking the direction of the ***positive*** gradient of $f(\theta_t)$