ECE 471/571 - Lecture 14
Gradient Descent

General Approach to Learning

- Optimization methods
  - Newton's method
  - Gradient descent
  - Exhaustive search through the solution space
- Objective functions
  - Maximum a-posteriori probability
  - Maximum likelihood estimate
  - Fisher's linear discriminant
  - Principal component analysis
  - k-nearest neighbor
  - Perceptron

Specify a model (objective function) and estimate its parameters
Use optimization methods to find the parameters
- $1^{st}$ derivative = 0
- Gradient descent
- Exhaustive search through the solution space
Newton-Raphson Method

**Used to find solution to equations**

According to Taylor series:
\[ f(x + \Delta x) = f(x) + \Delta x f'(x) \]
\[ f(x) + \Delta x f'(x) = 0 \Rightarrow \Delta x = -\frac{f(x)}{f'(x)} \]
\[ \Rightarrow x^{k+1} = x^k - \frac{f(x)}{f'(x)} \]

---

**Newton-Raphson Method vs. Gradient Descent**

- **Newton-Raphson method**
  - Used to find solution to equations
  - Find \( x \) for \( f(x) = 0 \)
  - The approach
    - Step 1: select initial \( x_0 \)
    - Step 2:
      \[ x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)} \]
    - Step 3: if \(|x^{k+1} - x^k| < \epsilon_1\), then stop; else \( x^k = x^{k+1} \) and go back step 2.

- **Gradient descent**
  - Used to find optima, i.e. solutions to derivatives
  - Find \( x^* \) such that \( f(x^*) < f(x) \)
  - The approach
    - Step 1: select initial \( x_0 \)
    - Step 2:
      \[ x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)} = x^k - \epsilon f'(x^k) \]
    - Step 3: if \(|x^{k+1} - x^k| < \epsilon_2\), then stop; else \( x^k = x^{k+1} \) and go back step 2.

---

**On the Learning Rate**

\[ x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)} = x^k - \epsilon f'(x^k) \]
Geometric Interpretation

Gradient of tangent is 2