Questions

• Need to have a big picture on the cost functions and the corresponding optimization approaches we’ve learned in this semester
• The analogy between Newton’s method and Gradient Descent
• What is the geometrical interpretation of GD?
• What is the physical meaning of the learning rate?
• Why does the learning rate need to be very small?
General Approach to Learning

- Specify a model (objective function) and estimate its parameters
- Learning algorithms
  - Maximum Posterior Probability
  - Maximum Likelihood estimate
  - Fisher’s linear discriminant
  - Principal component analysis
  - k-nearest neighbor
  - Kmeans
  - Logistic regression
- Use optimization methods to find the parameters
  - Exhaustive search through the solution space
  - 1st derivative = 0, Newton-Raphson Method
  - Gradient descent
Newton-Raphson Method

Used to find solution to equations

According to Taylor series: \( f(x + \Delta x) \approx f(x) + \Delta x f'(x) \)

\[ f(x) + \Delta x f'(x) = 0 \Rightarrow \Delta x = -\frac{f(x)}{f'(x)} \]

\[ \Rightarrow x^{k+1} = x^k - \frac{f(x)}{f'(x)} \]
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 \( |x^{k+1} - x^k| < \varepsilon \), 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:
      \[
      x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)} = x^k - cf'(x^k)
      \]
    - Step 3: if \( |x^{k+1} - x^k| < \varepsilon \), then stop; else \( x^k = x^{k+1} \) and go back step 2.

\[
\begin{align*}
  f(x) &= x^2 - 5x - 4 \\
  f(x) &= xc\cos(x)
\end{align*}
\]
On the learning rate

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

Gradient of tangent is 2