Lecture 3
How to minimize cost

Sung Kim <hunkim+mr@gmail.com>
Acknowledgement

- Andrew Ng’s ML class
  - https://class.coursera.org/ml-003/lecture
  - http://www.holehouse.org/mlclass/ (note)

- Convolutional Neural Networks for Visual Recognition.
  - http://cs231n.github.io/

- Tensorflow
  - https://www.tensorflow.org
  - https://github.com/aymericdamien/TensorFlow-Examples
Hypothesis and Cost

\[ H(x) = Wx + b \]

\[ \text{cost}(W, b) = \frac{1}{m} \sum_{i=1}^{m} (H(x^{(i)}) - y^{(i)})^2 \]
Simplified hypothesis

\[ H(x) = Wx \]

\[ cost(W) = \frac{1}{m} \sum_{i=1}^{m} (Wx^{(i)} - y^{(i)})^2 \]
What $cost(W)$ looks like?

$$cost(W) = \frac{1}{m} \sum_{i=1}^{m} (W x^{(i)} - y^{(i)})^2$$

<table>
<thead>
<tr>
<th>x</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

- $W=1$, $cost(W)=?$
What \( \text{cost}(W) \) looks like?

\[
\text{cost}(W) = \frac{1}{m} \sum_{i=1}^{m} (W x^{(i)} - y^{(i)})^2
\]

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y )</th>
<th>( \text{cost}(W) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4.67</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>?</td>
</tr>
</tbody>
</table>

- \( W=1 \), \( \text{cost}(W)=0 \)
  \[
  \frac{1}{3} \left( (1 \times 1 - 1)^2 + (1 \times 2 - 2)^2 + (1 \times 3 - 3)^2 \right)
  \]

- \( W=0 \), \( \text{cost}(W)=4.67 \)
  \[
  \frac{1}{3} \left( (0 \times 1 - 1)^2 + (0 \times 2 - 2)^2 + (0 \times 3 - 3)^2 \right)
  \]

- \( W=2 \), \( \text{cost}(W)=? \)
What \(\text{cost}(W)\) looks like?

- \(W=1\), \(\text{cost}(W)=0\)
- \(W=0\), \(\text{cost}(W)=4.67\)
- \(W=2\), \(\text{cost}(W)=4.67\)
What \textit{cost}(W) looks like?

\[
\text{cost}(W) = \frac{1}{m} \sum_{i=1}^{m} (W x^{(i)} - y^{(i)})^2
\]
How to minimize cost?

\[ \text{cost}(W) = \frac{1}{m} \sum_{i=1}^{m} (Wx^{(i)} - y^{(i)})^2 \]
Gradient descent algorithm

- Minimize cost function
- Gradient descent is used in many minimization problems
- For a given cost function, $cost (W, b)$, it will find $W, b$ to minimize cost
- It can be applied to more general function: $cost (w1, w2, \ldots)$
How it works?
How would you find the lowest point?
How it works?

- Start with initial guesses
  - Start at 0,0 (or any other value)
  - Keeping changing $W$ and $b$ a little bit to try and reduce cost($W, b$)

- Each time you change the parameters, you select the gradient which reduces cost($W, b$) the most possible

- Repeat

- Do so until you converge to a local minimum

- Has an interesting property
  - Where you start can determine which minimum you end up

http://www.holehouse.org/mlclass/01_02_Introduction_regression_analysis_and_gr.html
Formal definition

\[
\text{cost}(W) = \frac{1}{m} \sum_{i=1}^{m} (W x^{(i)} - y^{(i)})^2
\]

\[
\text{cost}(W) = \frac{1}{2m} \sum_{i=1}^{m} (W x^{(i)} - y^{(i)})^2
\]
Formal definition

\[ cost(W) = \frac{1}{2m} \sum_{i=1}^{m} (W x^{(i)} - y^{(i)})^2 \]

\[ W := W - \alpha \frac{\partial}{\partial W} cost(W) \]
Formal definition

\[ W := W - \alpha \frac{\partial}{\partial W} \frac{1}{2m} \sum_{i=1}^{m} (W x^{(i)} - y^{(i)})^2 \]

\[ W := W - \alpha \frac{1}{2m} \sum_{i=1}^{m} 2(W x^{(i)} - y^{(i)}) x^{(i)} \]

\[ W := W - \alpha \frac{1}{m} \sum_{i=1}^{m} (W x^{(i)} - y^{(i)}) x^{(i)} \]
Calculate the Derivative of ...

\((x+y)^2\)

This will be calculated:

\[
\frac{d}{dx} \left[(x-a) - y\right]^2
\]

Not what you mean? Use parentheses! Set differentiation variable and order in "Options".

Done! See the result further below.
Gradient descent algorithm

\[ W := W - \alpha \frac{1}{m} \sum_{i=1}^{m} (Wx^{(i)} - y^{(i)})x^{(i)} \]
Convex function

www.holehouse.org/mlclass/
convex function

$$cost(W, b) = \frac{1}{m} \sum_{i=1}^{m} (H(x^{(i)}) - y^{(i)})^2$$
Next
Multivariable logistic regression