The Linear Model and Gradient Descent
2018-08-23

The Simple Linear Model

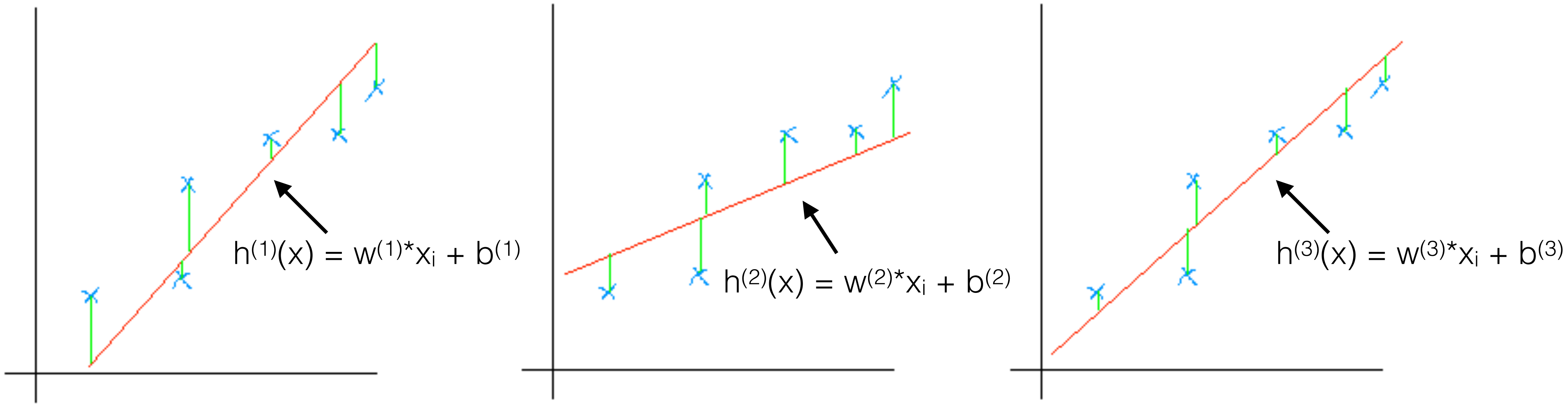
h(x) = w*x + b

-x is a feature or independent variable

-h(x), the hypothesis, is a predicted value for the target or output variable

-w and b, sometimes called the weight and bias, are constants

Error



- For each prediction $h(x_i)$ there will be an error e_i, where $e_i = h(x_i) - y_i$
- In the above examples:
  - Blue crosses are actual data points
  - Each of the red lines represents a linear model
  - Vertical green bars are the differences between the data (actual experience y_i) and the model (predicted outcome $h(x_i)$)

Cost Function

- We want some function J(w, b) that informs us how well we have chosen w and b
- Consider Mean squared error:

$$J(w,b) = \frac{1}{2n}\sum_{i=1}^{n} {e_i}^2$$

$$J(w,b) = \frac{1}{2n}\sum_{i=1}^{n} (h(x_i) - y_i)^2$$

$$J(w,b) = \frac{1}{2n}\sum_{i=1}^{n} (w * x_i + b - y_i)^2$$

Optimizing The Cost Function

Calculate the partial derivatives of cost w.r.t **w** and **b**:

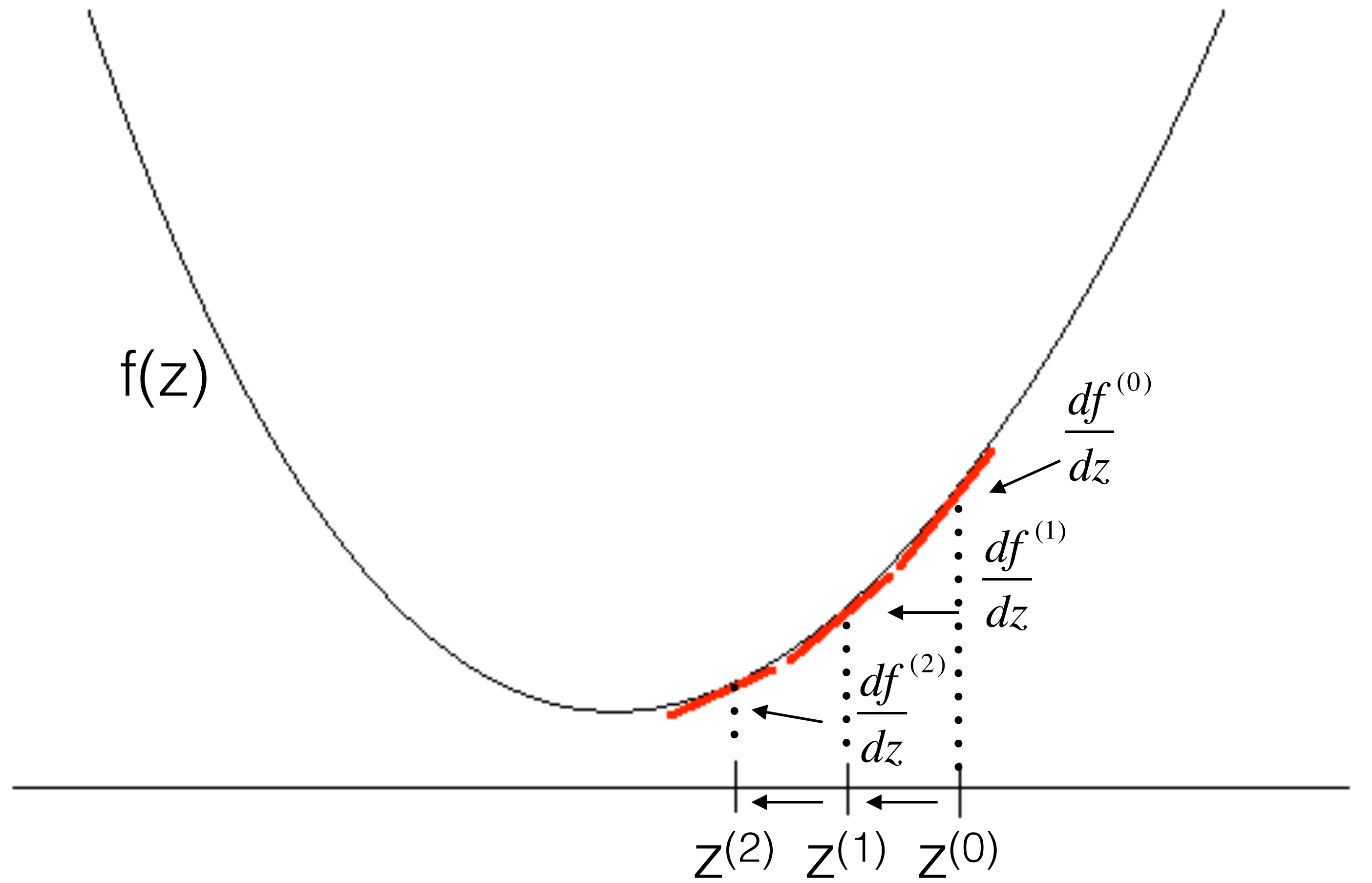
$$J(w,b) = \frac{1}{2n}\sum_{i=1}^{n}(w * x_i + b - y_i)^2$$

$$\frac{\partial J}{\partial w} = \frac{1}{n}\sum_{i=1}^{n} x_i(w * x_i + b - y_i)$$

$$\frac{\partial J}{\partial b} = \frac{1}{n}\sum_{i=1}^{n}(w * x_i + b - y_i)$$

Gradient Descent

Goal: Find the value of z that minimizes f(z)



- Choose an initial value, $z^{(0)}$ arbitrarily
- Update: $z^{(t+1)} = z^{(t)} - \alpha \frac{df^{(t)}}{dz}$
- Continue until $|f(z)^{(t+1)} - f(z)^{(t)}|$ is less than a predetermined threshold

Summary

- Goal: Construct a model of the form h(x) = w*x + b
- Define a function $J(w,b)=\frac{1}{2n}\sum_{i=1}^{n}(h(x_i)-y_i)^2$ to evaluate the performance of our model
- Find w and b that minimize J
- Use gradient descent to find optimal values of w and b

Linear Regression With Multiple Features

- Instead of having one feature or input x, we now have a vector of inputs (x_1, x_2, …, x_m)

- Our hypothesis becomes $h(x) = w_1x_1 + w_2x_2 + \ldots + w_mx_m + b$

- Cost becomes $J(w_1, w_2, ..., w_m, b) = \frac{1}{2n}\sum_{i=1}^{n}(w_1 * x_{i,1} + w_2 * x_{i,2} + ... + w_m * x_{i,m} + b - y_i)^2$

- The partial derivative of J w.r.t. wk is:

$$\frac{\partial J}{\partial w_k} = \frac{1}{n}\sum_{i=1}^{n} x_{i,k}(w_1 * x_{i,1} + w_2 * x_{i,2} + ... + w_m * x_{i,m} + b - y_i)$$

- The update step for gradient descent becomes:

$$(z_1, z_2, \ldots, z_j)^{(t+1)} = (z_1, z_2, \ldots, z_j)^{(t)} - \alpha\left(\frac{\partial f}{dz_1}, \frac{\partial f}{dz_2}, ..., \frac{\partial f}{dz_j}\right)^{(t)}$$

Additional Resources

- sklearn library for algorithms and datasets: http://scikit-learn.org/stable/datasets/index.html
- Continuing education:
    - Beginner: https://www.coursera.org/learn/machine-learning
    - Advanced: https://www.coursera.org/specializations/deep-learning
- MNIST: http://yann.lecun.com/exdb/mnist/