Week 2

Feature scaling / Mean Normalization

- Input values for gradient descent should be in roughly the same range, ideally something like: $-1 <= x_1 <= 1$
- Feature scaling: divide the input values by the range (max - min) or std_dev (guarantees distribution has a range of 1)
- Mean Normalization: subtract the mean from the observation, resulting in a dsitribution centered on 0
- Putting it together: $x_i := \frac{x_1 - \mu_i}{s_i}$ (where s_i is the std_dev or the range)

Learning Rate

- Learning rate is the parameter α
- Debugging gradient descent: if the cost function $J(\theta)$ ever increases over time, the learning rate is probably too large
- Automatic convergence test: declare convergence if $J(\theta)$ ever decreases by less than ϵ in an iteration

Normal Equation

- Can be used to compute the minimum for the cost function analytically
- Equation: $\theta = (X^T X)^{-1} X^T y$
- Runtime $O(n^3)$ and requires matrix inversion, will be slow if n is large
- Does not require α or a learning rate