ML notes from Stanford University Course

From the basics ML can be classified as unsupervised learning and supervised learning. Supervised Learning tasks use labeled data fed into a model to output a prediction on new data

Unsupervised Learning tasks use unlabeled data and try to find clusters or groups of data that indicate some form of a trending.

A Regression is a model that predicts an continuous value.

A classification distinguishes data into discrete classes.

Univariate Linear Regression:

Y = B0 + B1x

Y as the variable to be predicted, x as the input variable

MultiVariate Linear Regression:

$$h_\theta(x) = \theta^T x = \theta_0 x_0 + \cdots + \theta_n x_n$$

Cost Function:
A measure of how accurate the model is at predicting a value.

$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(\mathrm{x}^{(\mathrm{i})}\right) - \mathrm{y}^{(i)}\right)^2$$

Gradient Descent :
Repeat{

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1, \dots, \theta_n)$$

} Simultaneous updates for each j = 0,…,n.
Ie the partial derivative of the cost function with respect to an I, where 0<=i<=n.

Regularization for Linear Regression:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)x_j{}^{(i)} + \frac{\lambda}{m}\theta_j$$

For Gradient Descent:

$$\theta_0 := \theta_0 - \alpha \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(\mathrm{x}^{(\mathrm{i})}\right) - \mathrm{y}^{(i)}\right) x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha\left(\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(\mathrm{x}^{(\mathrm{i})}\right) - \mathrm{y}^{(i)}\right) x_0^{(i)} + \frac{\lambda}{m}\theta_j\right)$$

Or by factorizing:

$$\theta_j := \theta_j(1 - \alpha\frac{\lambda}{m}) - \alpha\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(\mathrm{x}^{(\mathrm{i})}\right) - \mathrm{y}^{(i)}\right) x_j^{(i)}$$

Regularization takes care of non-invertibility of normal eqn for linear regression.
If $\lambda > 0$:

$$\theta = (\boldsymbol{X^T X} + \boldsymbol{\lambda}\begin{bmatrix}\mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{1}\end{bmatrix})^{-\mathbf{1}} X^T y$$

In bold will always be invertible.