MIT Machine Learning notes

Vincent Courtois

Linear Classifiers

Perceptron Algorithm $\theta = \theta_0 = 0$. if $y_i(\theta.x_i) \leq 0$ then $\theta = \theta + y_i x_i, \theta_0 = \theta_0 + y_i$
Large Margin Optimization
signed distance to boundary $\frac{y^{(i)}(\theta.x^{(i)}+\theta_0)}{||\theta||}$. Margins are $\frac{1}{||\theta||}$ away from boundary.
hinge loss $Loss_h(y^{(i)}(\theta.x^{(i)} + \theta_0)) = Loss_h(z) = 0$ if z>=1, $1 - z$ if z<1
objective function $J(\theta, \theta_0) = \frac{1}{n}\Sigma_{i=1}^{n} Loss_h(y^{(i)}(\theta.x^{(i)} + \theta_0)) + \frac{\lambda}{2}||\theta||^2$

Optimization algorithms

Stochastic gradient descent $\theta \leftarrow \theta - \eta_t \nabla_\theta [Loss_h(y^{(i)}\theta.x^{(i)}) + \frac{\lambda}{2}||\theta||^2]$