1 Least Square Regression Problem

$h_\theta(x) = \theta^T x$

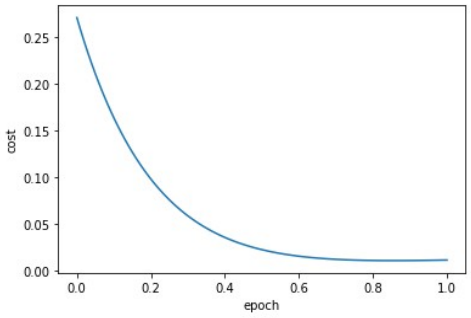
1.1 Batch Gradient Decent

cost function: $\frac{1}{2m}\sum_{i=0}^{m-1}(h_\theta(X^i) - y^i)^2$

Update Rule: m = 600
$for\ j = 0....7:$
$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=0}^{m-1}(h_\theta(X^i) - y^i)X_j^i$

1.1.1 Result Plot

epochs = 100, $\alpha = 0.0001$



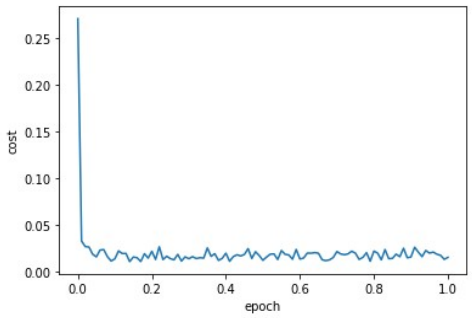
1.2 Stochastic Gradient Decent

cost function: $\frac{1}{2}(h_\theta(X^i) - y^i)^2$

Update Rule: m = 600
$for\ i = 0....599:$
$\quad for\ j = 0....7:$
$\quad\quad \theta_j := \theta_j - \alpha(h_\theta(X^i) - y^i)X_j^i$

1.2.1 Result Plot

epochs = 100, $\alpha = 0.00001$



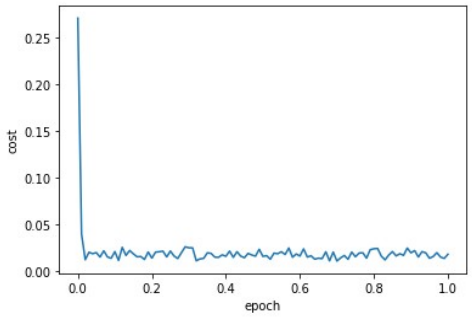
1.3 Mini Batch Gradient Decent

cost function: $\frac{1}{k}\sum_{i=0}^{k-1}(h_\theta(X^k) - y^k)^2$

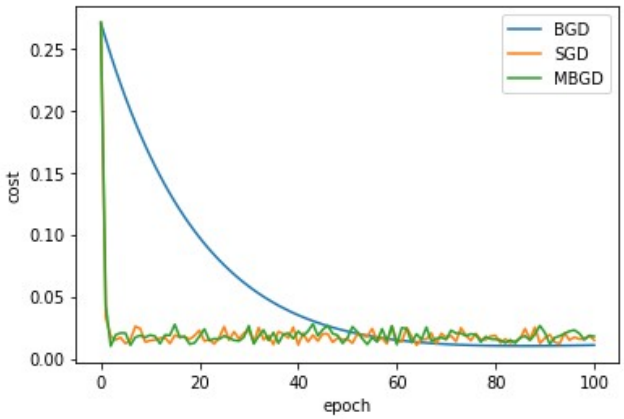
Update Rule: k = 10, m = 600
$for\ i = 0..9..19....599:$
$\quad for\ j = 0....7:$
$\quad\quad \theta_j := \theta_j - \alpha\frac{1}{b}\sum_{k=i}^{(i+k)-1}(h_\theta(X^k) - y^k)X_j^k$

1.3.1 Result Plot

epochs = 100, $\alpha = 0.00001$, b = 10



1.4 BGD vs SGD vs MBGD



2 Classification Problem

$h_\theta(x) = \frac{1}{1+e^{\theta^T x}}$

2.1 Batch Gradient Decent

cost function: $-\frac{1}{m}\sum_{i=0}^{m-1}(\ y^i log(h_\theta(X^i)) + (1-y^i) log(1-h_\theta(X^i))\ )$

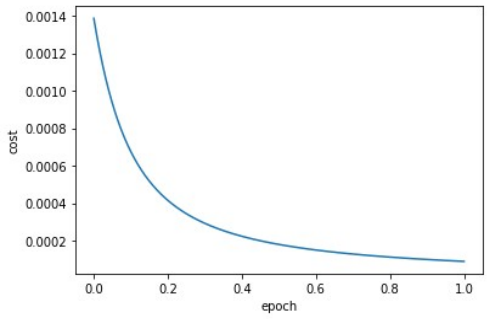
Update Rule: m = 600
$for\ j = 0....7:$
$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=0}^{m-1}(h_\theta(X^i) - y^i)X_j^i$

2.1.1 Accuracy

0.806

2.1.2 Result Plot

epochs = 1000, $\alpha = 0.0001$



2.2 Stochastic Gradient Decent

cost function: $-(y^i log(h_\theta(X^i)) + (1-y^i) log(1-h_\theta(X^i)))$

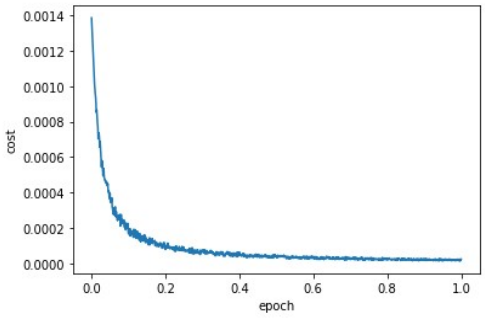
Update Rule: m = 600
$for\ i = 0....599:$
$\quad for\ j = 0....7:$
$\quad\quad \theta_j := \theta_j - \alpha(h_\theta(X^i) - y^i)X_j^i$

2.2.1 Accuracy

0.806

2.2.2 Result Plot

epochs = 1000, $\alpha = 0.000001$



2.3 Mini Batch Gradient Decent

cost function: $-\frac{1}{k}\sum_{i=0}^{k-1}(\ y^k log(h_\theta(X^k)) + (1-y^k) log(1-h_\theta(X^k))\ )$

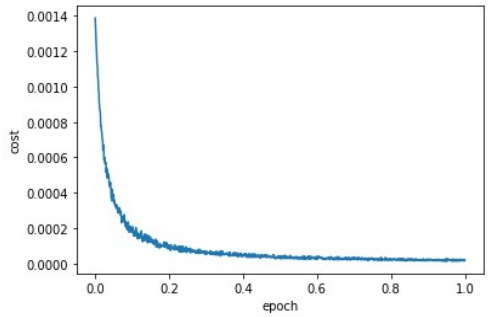
Update Rule: k = 10, m = 600
$for\ i = 0..9..19....599:$
$\quad for\ j = 0....7:$
$\quad\quad \theta_j := \theta_j - \alpha\frac{1}{b}\sum_{k=i}^{(i+k)-1}(h_\theta(X^k) - y^k)X_j^k$

2.3.1 Accuracy

0.806

2.3.2 Result Plot

epochs = 1000, $\alpha = 0.000001$, b = 10



2.4 BGD vs SGD vs MBGD