

Machine Learning Dictionary

Optimization

Gradient descent

Stochastic gradient descent

```
theta = theta - lr*grad(loss(theta))
```

Momentum

Adam

Gradient accumulation