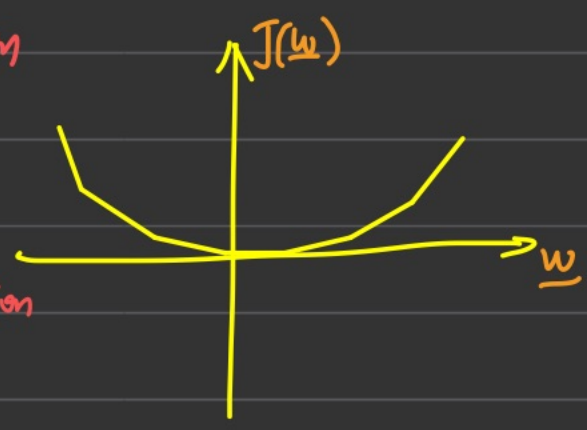
Perceptron Criterion Function

$$J(\underline{w}) = -\sum_{n=1}^{N} [\![g(z_n \underline{x}_n) \le 0]\!] \underline{w}^T z_n \underline{x}_n$$

indicator function

augmented notation



Minimize it by gradient descent (GD)

$$\therefore \nabla_{\underline{w}} J(\underline{w}) = -\sum_{n=1}^{N} [\![g(z_n \underline{x}) \le 0]\!] z_n \underline{x}_n$$

$$\underline{w}(i+1) = \underline{w}(i) - \eta(i) \nabla_{\underline{w}} J(\underline{w}) = \underline{w}(i) + \eta(i) \sum_{n=1}^{N} [\![g(z_n \underline{x}) \le 0]\!] z_n \underline{x}_n$$

where $\eta(i) > 0 \ \forall i$

stop when $\nabla_{\underline{w}} J(\underline{w}) = 0$ if linearly separable
or reach maximum iteration limit

← batch GD

Gradient Descent (GD) Algorithm

- Batch GD

  For each epoch: $\underline{w}(i+1) = \underline{w}(i) - \eta(i) \nabla_{\underline{w}} J(\underline{w})$

- Sequential GD

$$J(\underline{w}) = \sum_{n=1}^{N} J_n(\underline{w})$$

For each epoch:

For each data point $\underline{x}_n$:

$$\underline{w}(i+1) = \underline{w}(i) - \eta(i) \nabla_{\underline{w}} J_n(\underline{w})$$

Stochastic GD - variant 1

For each epoch:

Randomly shuffle dataset

For each data point x_n:

$$\underline{w}(i+1) = \underline{w}(i) - \eta(i) \nabla_{\underline{w}} J_n(\underline{w})$$

Stochastic GD - variant 2

For each epoch:

For each iteration:

Randomly pick one data point x_n

$$\underline{w}(i+1) = \underline{w}(i) - \eta(i) \nabla_{\underline{w}} J_n(\underline{w})$$

Mini-Batch GD

For each epoch:

Split dataset into batches

For each batch:

$$\underline{w}(i+1) = \underline{w}(i) - \eta(i) \nabla_{\underline{w}} J(\underline{w})$$

only use data in batch

or

For each iteration:

Randomly pick m data

$$\underline{w}(i+1) = \underline{w}(i) - \eta(i) \nabla_{\underline{w}} J(\underline{w})$$

only use picked data