Gradient Descent

Prof. Dr. Wawan Setiawan, M.Kom.
Yaya Wihardi, S.Kom., M.Kom.
Email: yayawihardi@upi.edu

Department of Computer Science Education

Universitas Pendidikan Indonesia

Training pada Perceptron

$$w_i \leftarrow w_i + \Delta w_i$$

dimana :

$$\Delta w_i = \eta\,(t - o)x_i$$

- $t = c(\vec{x})$ adalah target value
- o adalah output dari perceptron
- η disebut learning rate,

  biasanya konstanta bernilai rendah (mis : 1)

Permasalahan Optimasi pada Perceptron

Perhatikan model linear berikut :

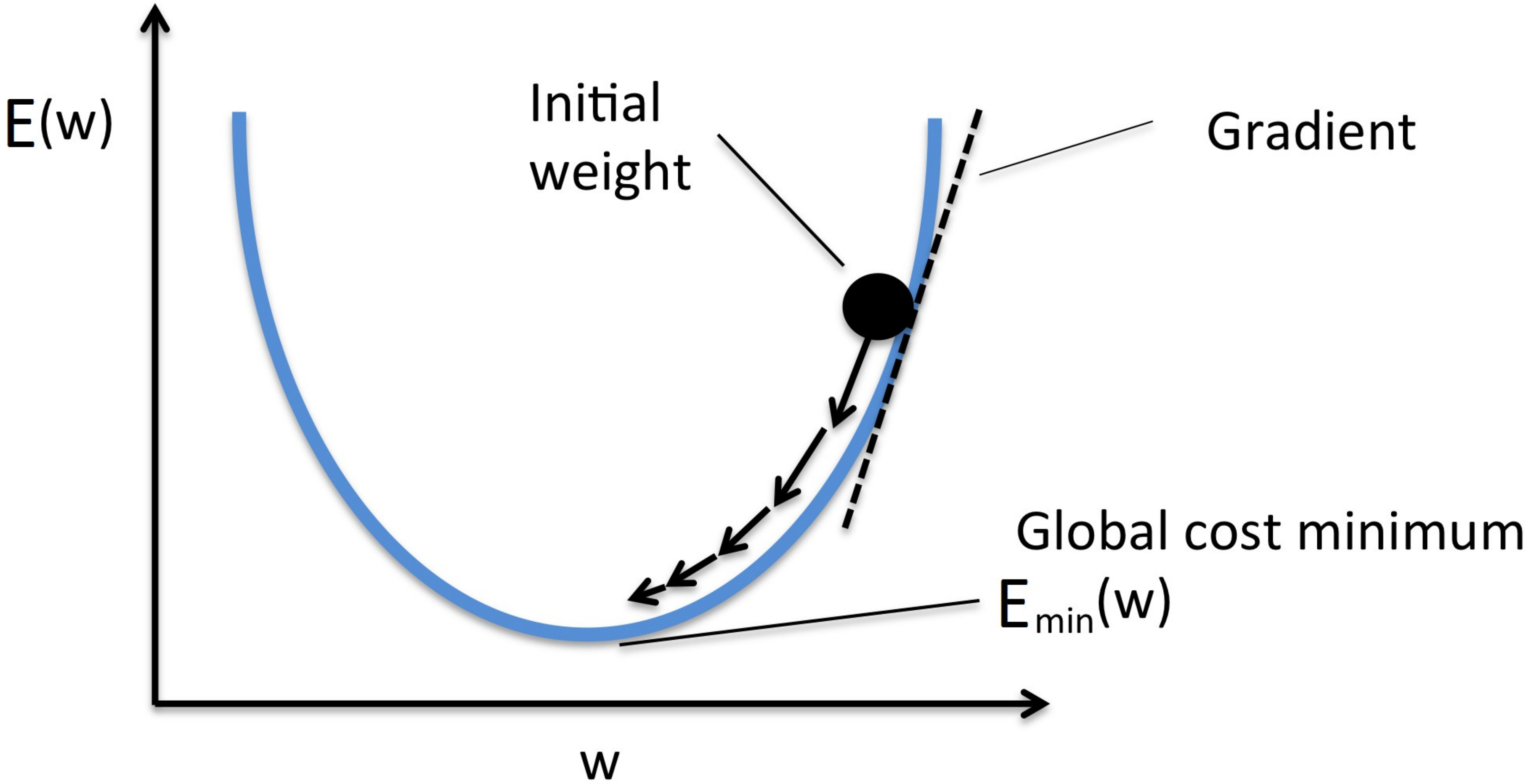
$$o = w_0 + w_1 x_1 + \ldots + w_n x_n$$

Ide : cari nilai w_i' yang meminimumkan fungsi error/loss berikut :

$$E[\vec{w}] \equiv \tfrac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

Dimana D merupakan set training sample

Gradient Descent



Aturan Pembelajaran

Gradient $\nabla E[\vec{w}] \equiv \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \ldots, \frac{\partial E}{\partial w_n} \right]$

Training rule : $\Delta w_i = -\eta \, \nabla E[\vec{w}]$

i.e., $\Delta w_i = -\eta \, \frac{\partial E}{\partial w_i}$

Menghitung Gradient

$$\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i} \frac{1}{2} \sum_d (t_d - o_d)^2$$

$$= \frac{1}{2} \sum_d \frac{\partial}{\partial w_i} (t_d - o_d)^2$$

$$= \frac{1}{2} \sum_d 2(t_d - o_d) \frac{\partial}{\partial w_i} (t_d - o_d)$$

$$= \sum_d (t_d - o_d) \frac{\partial}{\partial w_i} (t_d - \vec{w} \cdot \vec{x}_d)$$

$$\frac{\partial E}{\partial w_i} = \sum_d (t_d - o_d)(-x_{i,d})$$

Gradient Descent

$\text{GRADIENT}-\text{DESCENT}(training\_examples, \eta)$

Each training examples is a pair of the form $<\vec{x}, t>$*, where* $\vec{x}$ *is the vector of input values and t is the target output value.* η *is the learning rate (e.g., .05).*

- Initialize each w_i to some small random value
- Until the termination condition is met, do

  - Initialize each Δw_i to zero.

  - For each $<\vec{x}, t>$ in $training\_examples$, do

    * Input the instance $\vec{x}$ and compute output o

    * For each linear unit weight w_i, do

      $$\Delta w_i \leftarrow \Delta w_i + \eta (t - o) x_i$$

  - For each linear unit weight w, do

    $$w_i \leftarrow w_i + \Delta w_i$$

Batch vs Stochastic (Incremental) GD

Batch mode Gradient Descent:
Do until satisfied:

1. Compute the gradient $\nabla E_D[\vec{w}]$
2. $\vec{w} \leftarrow \vec{w} - \eta \nabla E_D[\vec{w}]$

$$E_D[\vec{w}] \equiv \tfrac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

Incremental mode Gradient Descent:
Do until satisfied:
- For each training example d in D

1. Compute the gradient $\nabla E_d[\vec{w}]$
2. $\vec{w} \leftarrow \vec{w} - \eta \nabla E_d[\vec{w}]$

$$E_d[\vec{w}] \equiv \tfrac{1}{2} (t_d - o_d)^2$$

Thank You