Topics in Computer Science Spring 2025
COS 470 Lecture 4

Objectives

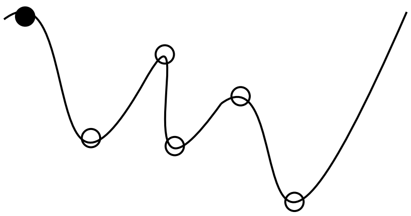
- Optimizers
- Gradient Descent

Optimizers

Goal: Find "test" = max, or min

Loss function: minimize Two methods

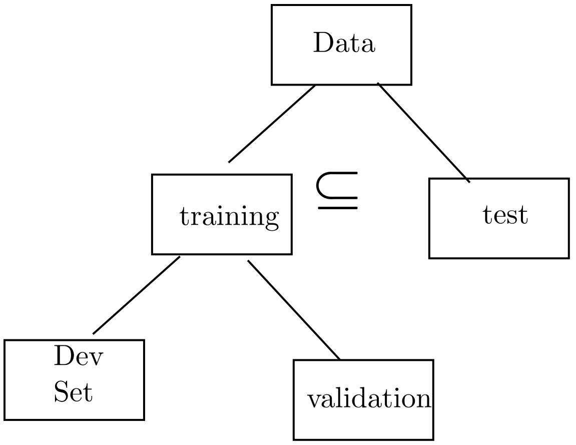
- exact take derivative = 0
- Gradient based method



initial value = seed the generator

1 vanilla gradient descent

update function: $\boxed{w_{t+1} = w_t - \gamma\nabla L(W)}$
L(W) is the loss function
γ is the learning rate, a hyperparameter tunable via cross validation



frequent issues:

- gets stuck at local minimum
- vanishing gradient
- if learning rate is increased can get unstuck but risk jumping over desired points if increased to much

2 momentum

intuitive: $w_{t+1} = w_t - \gamma\nabla L)w_t +$ (past gradients)
Actual: $\boxed{v_0 = 0}$
$\boxed{v_{t_1} = \xi v_t + \nabla L(w_t)}$
$\boxed{w_{t_1} = w_t - \gamma v_{t+1}}$
issue: can move past desired point because of momentum of past gradients
evaluating performance: $\boxed{\text{Count number of iterations}}$

3 Gradient-based Nesterov

$\boxed{v_0 = 0}$
$\boxed{v_{t+1} = \xi v_t \nabla L(w_t)}$
$\boxed{w_{t+1} = w_t - \gamma(\delta v_{t+1} + \nabla L(w_t))}$
look ahead : δv_{t+1}