eCornell

CHEAT SHEET

Gradient Descent

Algorithm Name	Gradient Descent
Description	Gradient Descent is a minimization method that uses only the gradient information. Essentially, you update the parameters by stepping in $-\nabla f$, which is the steepest decreasing direction for function f.
Applicability	Minimization problems.
Assumptions	The objective function has to be differentiable, namely, the gradient exists.
Underlying Mathematical Principles	• Gradient • Partial derivatives
Additional Details	• Gradient Descent gives optimal solution if the loss function is convex. • If loss function is not convex, gradient descent might only produce a local minimum rather than the global minimum. • The learning rate α is a hyperparameter (the model does not learn it but you have to manually tune it).
Example	You can use gradient descent to find the optimal weight vector for the logistic loss function, the optimal minimum solution of which has no closed-form expression.