Recitation Week 14

Ashwin Bhola

CDS, NYU

Dec 4^{th}, 2019

Gradient Descent

1. Given N training points $\{(x_k, y_k)\}_{k=1}^N$, $x_k \in \mathbb{R}^d$, $y_k \in \{-1,1\}$, we seek a linear discriminant function $f(x) = w^T x$
    1. Let $S = \{x \in \mathbb{R}^d | w^T x = a\}$. What's the relation between $w \in \mathbb{R}^d$ and S?
    2. Suggest a method for determining if the points are perfectly linearly separable
    3. Let's say we decide to take the loss function as the exponential loss, $L(z) = e^{-z}$, where z is defined to be the margin i.e. $z = yf(x)$, y being the labels of x. Derive the update rule for batch gradient descent for exponential loss. What's the computational cost of this update in terms of N and d?
    4. Can we express loss for logistic regression as a function of z?
    5. The step size in gradient descent is not actual "step size" because the true length of the step in gradient descent will be $\alpha||\nabla_w J(w)||$. Someone suggests that we should scale the gradients to make them unit norm before we do gradient descent. What's the problem with this?
    6. What are the different possible stopping criteria for gradient descent algorithm?

Optimality conditions

1. Consider f to be twice differentiable function. True or False:
    1. Convex functions do not have saddle points
    2. For x to be a local minimum of a function $f(x)$, $\nabla f(x) = 0 \; and \; \mathrm{H}_f(x) \succ 0$
    3. The necessary and sufficient condition for local optimality in unconstrained convex optimization is $\nabla f(x) = 0$
    4. The necessary and sufficient condition for local optimality in unconstrained optimization is $\nabla f(x) = 0$
    5. For convex functions, the direction of steepest descent is same as the direction towards global optima
    6. For a convex function, the first order Taylor approximation at any point is a global under estimator of the function
    7. $e^x \geq 1 + x$
2. Consider the function $f(x) = x^T A x + b^T x + c$, where $A \in \mathbb{R}^{n \times n}$ is symmetric, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$. Suppose we are trying to minimize $f(x)$. Find the minimum value of f(x) when (i) A is positive definite (ii) A has a strictly negative eigenvalue