CS 189 Introduction to Machine Learning
Fall 2017

DIS6

1 Gradient descent for simple functions

(a) Recall Taylor's theorem for twice differentiable functions of vectors, which holds for all $x, y \in \mathbb{R}^d$:

$$f(y) = f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(\widetilde{x})(y-x),$$

for some $\widetilde{x}$. Show that the function f is convex if $\nabla^2 f(x)$ is positive semidefinite for all $x \in \mathbb{R}^d$.

(b) Let $L \geq 0$. Consider the function of one variable $f(x) = \frac{L}{2}x^2$. Show that it is convex.

(c) Derive the gradient descent update where we use a step-size of γ and start at some point $x^{(0)} \neq 0$.

(d) What does the behavior look like for the above setting and the choices $\gamma \in \{1/L, 2/L\}$?

(e) Consider the above setup and assume we use a step size $\gamma \in \left[0, \frac{2}{L}\right)$. Also assume that $\gamma \neq 1/L$. How many steps does it take for us to converge to within ε of the optimum (as a function of the tuple $(\gamma, L, |x^{(0)}|, \varepsilon)$)?

(f) How do your answers above change if $f(x) = \frac{L}{2}(x-c)^2$ for some constant c?

(g) Let $L \geq m \geq 0$. Now consider the function of two variables $f(x) = \frac{L}{2}x_1^2 + \frac{m}{2}x_2^2$. Show that the function is convex by computing its Hessian $\nabla^2 f(x)$.

(h) With the setup of the previous part, let us say we started at the point $(0,5)$. What is the maximum step-size that results in convergence? How would your answer change if we started at the point $(5,0)$?

(i) Derive closed form expressions for the iterations if we start at the point (a,b), and run gradient descent with step-size γ. Start by writing out the result of the first iteration as $A\begin{bmatrix} a \\ b \end{bmatrix}$ for some matrix A.

(j) Now consider the function of one variable $f(x) = L|x|$. Is this function convex? Discuss how performing gradient descent with a fixed step-size performs on this function.