ST510: Foundations of Machine Learning – Week 2

Convex Optimisation (theory)

1. Let f be a convex differentiable function. Prove that any stationary point x^* is a global minimiser of f.

2. Let $f : \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ be convex functions. Prove that $f + g$ is also convex. How about $f - g$? And fg?

3. Recall that a set S is convex if for any $\boldsymbol{x}, \boldsymbol{y} \in S$, one also has $t\boldsymbol{x} + (1-t)\boldsymbol{y} \in S$ for any $t \in (0, 1)$. Let f be a convex function. Prove that the set of global minimizers of f is a convex set.

4. Let $\{x_k\}$ be a sequence in $\mathbb{R}$ that converges to x^*.

   The convergence is linear if there is a constant $\rho \in (0, 1)$ such that

   $$\frac{|x^{k+1} - x^*|}{|x^k - x^*|} \leq \rho, \text{ for all sufficiently large } k.$$

   The convergence is quadratic if there is a positive constant C such that

   $$\frac{|x^{k+1} - x^*|}{|x^k - x^*|^2} \leq C, \text{ for all sufficiently large } k.$$

   Determine the convergence rate of the following sequences:

   (a) $x^k = 1/k$

   (b) $x^k = 1 + (0.5)^{2^k}$

   (c) $x^k = 1/(k!)$

5. Consider the following two functions: $f(x) = |x^2 - 2|$ and $g(x) = (x^2 - 2)^2$. Apply Newton's method to find a (positive) minimiser of f and g. Does it work? Verify your conclusion in Python.

6. Consider the problem of minimising $f(\boldsymbol{x}) := f(x_1, x_2) = (2x_1^2 - x_2)^2 + 3x_1^2 - x_2$. Let $\boldsymbol{x}^0 := (1/2, 5/4)^T$.

   (a) Is the function convex?

   (b) Determine all the descent directions of f at $\boldsymbol{x}^0$.

   (c) What is the steepest descent direction?

   (d) Perform one iteration of the steepest descent method using an exact linear search. What is $\boldsymbol{x}^1$?

   (e) Will it converge to a global optimum?

7. Consider the problem of minimising $f(\boldsymbol{x}) := f(x_1, x_2) = (x_1 + 2x_2 - 3)^2 + (x_1 - 2)^2$. Let $\boldsymbol{x}^0 := (0, 0)^T$.

(a) Perform one iteration of Newton's method with an exact line search.

(b) Are there any descent directions from $\boldsymbol{x}^1$? Is $\boldsymbol{x}^1$ optimal?

(c) Does the starting point $\boldsymbol{x}^0$ matter here? Why/why not?

8. For coordinate descent algorithm, prove or disprove:

   (a) Given convex, differentiable f, if we are at a point $\boldsymbol{x}$ such that $f(\boldsymbol{x})$ is minimized along each coordinate axis, have we found a global minimizer?

   (b) What if f is convex but non-differentiable?

   (c) What if f is of the form $f(\boldsymbol{x}) = g(\boldsymbol{x}) + h_1(x_1) + \cdots + h_d(x_d)$, where g is convex and smooth and each h_i is convex for $i = 1, \ldots d$?