15-451/651 Algorithms, Spring 2019
Recitation #10 Worksheet

Multiplicative Weights

1. In lecture we saw that the simple procedure that multiplied the weight of each expert by $\frac{1}{2}$ whenever the expert made a mistake, resulted in
$$m = \text{ \#mistakes of algorithm } \leq 2.41(M + \log_2 n),$$
where M = #mistakes made by the best expert and n = # of experts. If we multiply the weight by 2/3 at each time, how does this analysis change?

2. In lecture, we got a mistake bound of $(1+\varepsilon)M + O(\frac{\ln n}{\varepsilon})$, using randomization. Let us now show that you cannot get better than $2M$ mistakes if you don't use randomness.

   There are two experts. One always predicts 0*. The other always predicts* 1*. Fix any deterministic algorithm* A *for prediction. Here is one sequence of days: each day, the actual outcome is the opposite of what the algorithm predicts.*

   After T days, the algorithm would have made T mistakes. Show that the better of the two experts makes at most $T/2$ mistakes. Hence infer that $m \geq 2M$.

Convex Functions (Recap)

Recall that a function f over $\mathbf{R}^n$ is *convex* if for any two inputs $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ and any $\lambda \in [0,1]$ we have
$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}).$$
In other words, the line segment from $(\mathbf{x}, f(\mathbf{x}))$ to $(\mathbf{y}, f(\mathbf{y}))$ stays "above" the function.

Moreover, recall that a set $K \subseteq \mathbf{R}^n$ is *convex* if for any two points $\mathbf{x}, \mathbf{y} \in K$ and any $\lambda \in [0,1]$ we have that the point $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \in K$. In other words, the line segment from $\mathbf{x}$ to $\mathbf{y}$ stays inside the set.

For all these statements below, it will help to draw the one-dimensional picture and stare at it.

4. The *epigraph* of a function $f : \mathbf{R}^n$ to $\mathbf{R}$ is the set $K := \{(\mathbf{x}, z) \mid \mathbf{x} \in \mathbf{R}^n, z \in \mathbf{R}, z \geq f(\mathbf{x})\} \subseteq \mathbf{R}^{n+1}$. Show that f is a convex function $\iff$ its epigraph is a convex set.

5. Show that level sets of a convex function are convex. I.e., for any z, the set $K_z = \{\mathbf{x} \in \mathbf{R}^n \mid f(\mathbf{x}) \leq z\}$ is a convex set. Show that the converse is not true — there are non-convex functions for which every K_z is convex.

6. A nice feature about convex functions is that any local minimum is a global minimum. Indeed, show that if $\mathbf{x}$ is not a global minimum, there is some direction in which the slope is negative at $\mathbf{x}$.

7. To formalize the notion of slope at a point $\mathbf{x}$, we consider the gradients. (Assume f is a differentiable function.) Then the gradient of f at $\mathbf{x}$ is given by

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right).$$

We can define a (differentiable) function f to be convex if for all $\mathbf{x}, \mathbf{y}$, we have

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle.$$

8. Consider the linear function $f(\mathbf{x}) = \sum_i c_i x_i = \langle \mathbf{c}, \mathbf{x} \rangle$. Show that f is convex, using both definitions. What is $\nabla f(\mathbf{x})$, the gradient of f at $\mathbf{x}$?

9. Do the same for the quadratic function $g(\mathbf{x}) = \sum_i c_i x_i^2$, where $c_i \geq 0$.

10. Suppose $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\intercal A \mathbf{x} + b\mathbf{x}$ for a symmetric matrix $A \in \mathbf{R}^{n \times n}$ and $b \in \mathbf{R}^n$. Compute the gradient $\nabla f(x)$ and the Hessian $\nabla^2 f(x)$. (Recall that the Hessian is an $n \times n$ matrix with i, j^{th} entry being $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$.) When is this function convex?