COS511 HW5

Pranjit Kalita

April 22, 2017

Ex. 1

Part a

In T rounds of a fair coin toss, if we have $|N_h - N_t| = k$, no. of combinations $= C(T, (T+k)/2)$
$E(|N_h - N_t|) = \sum_{k=0}^{T} 2 * k * C(T, (T+k)/2)/(2^T)$
$\Rightarrow$E(|N$_h - N_t|) = 1/(2^{T-1}) * \sum_{k=0}^{T} k * C(T, (T+k)/2)$
$\Rightarrow$E(|N$_h - N_t|) = 1/(2^{T-1}) * \sum_{k=0}^{T} k * T!/((T+k)/2)!((T-k)/2)!$
Now, $k = (T+k)/2 - (T-k)/2$
Substituting above,
$\Rightarrow$E(|N$_h - N_t|) = 1/(2^{T-1}) * \sum_{k=0}^{T}((T+k)/2) * T!/((T+k)/2)!((T-k)/2)! - ((T-k)/2)) * T!/((T+k)/2)!((T-k)/2)!$
Simplifying the above equation (too tedious to show calculation on Latex)
$\Rightarrow$E(|N$_h - N_t|) = 1/(2^{T-1}) * \sum_{k=0}^{T} T!/(((T+k)/2)-1)!((T-k)/2)! - T!/(((T-k)/2)-1)!((T+k)/2)!$
$\Rightarrow$E(|N$_h - N_t|) = T/(2^{T-1}) * \sum_{k=0}^{T}(T-1)!/(((T+k)/2)-1)!((T-k)/2)! - (T-1)!/(((T-k)/2)-1)!((T+k)/2)!$
$\Rightarrow$E(|N$_h - N_t|) = T/(2^{T-1}) * \sum_{k=0}^{T} C((T-1), (((T+k)/2)-1) - C((T-1), ((T+k)/2)$
Canceling out terms, $\Rightarrow$E(|N$_h - N_t|) = T/(2^{T-1}) * [C((T-1), (T/2-1)) - C((T-1), T)]$
$\Rightarrow$E(|N$_h - N_t|) = T/(2^{T-1}) * C((T-1), (T/2-1))$ (since second term is 0)
$\Rightarrow$E(|N$_h - N_t|) = T/(2^{T-1}) * T!/((T/2)!)^2$
By using Sterling's Approximation,
$\Rightarrow$E(|N$_h - N_t|) = (2/\pi) * \sqrt{2\pi T}$
$\Rightarrow$E(|N$_h - N_t|) = \Theta(\sqrt{T})$

Part b

Since at each toss of the coin, the learner incurs a loss of 0 or 1, and there are T total coin tosses, hence the expected loss of the learner $= \dfrac{T}{2}$.
Now, let X_t be a random variable, with $X_t = \{0, 1\}$ depicting the

incurred loss by the learner at each coin toss $t \in \{1, T\}$.
Thus, if one expert learner has X_t, the other has $(1 - X_t)$ at each coin toss $t \in \{1, T\}$.
Thus, for the learner, the loss will be of the form - $\{X_t, (1 - X_t)\}$, $X_t = \{0, 1\}$, $t \in \{1, T\}$.
Minimum expected loss of expert = $E[min\{\sum_{i=1}^{T} X_t, \sum_{i=1}^{T}(1 - X_t)\}]$
$= E[\frac{\sum_{i=1}^{T} X_t + \sum_{i=1}^{T}(1 - X_t)}{2} - \frac{|\sum_{i=1}^{T} X_t - \sum_{i=1}^{T}(1 - X_t)|}{2}]$
$= \frac{E}{2}[(\sum_{i=1}^{T} X_t + \sum_{i=1}^{T}(1 - X_t))] - \frac{E}{2}[|\sum_{i=1}^{T} X_t - \sum_{i=1}^{T}(1 - X_t)|]$
$= \frac{T}{2} - \frac{E}{2}[|\sum_{i=1}^{T} X_t - \sum_{i=1}^{T}(1 - X_t)|]$
Now, the second term can be related to **Part a** if $\sum_{i=1}^{T} X_t = N_h$, $\sum_{i=1}^{T}(1 - X_t) = N_t$, which then becomes similar to $E[—N_h - N_t|]$.
Thus, Minimum expected loss of expert = $\frac{T}{2} - \frac{\Theta\sqrt{T}}{2}$
$= \frac{T}{2} - \Theta\sqrt{T}$
Thus, the expected regret matches the upper bound for hedge $\Theta\sqrt{T}$.

Ex. 6

Let B be the outer algorithm, and A be the inner algorithm. We see that each iteration of A, for a value of m, runs:
$2^{m+1} - 2^m = 2^m$ times.
Thus, $Regret = O(\alpha\sqrt{T}) = O(\alpha\sqrt{2^m})$
Now, for finding B's regret, we will have to sum over all A's as:
$Regret_B = \sum_{m=1}^{logT} O(\alpha\sqrt{2^m})$
$\Rightarrow$Regret$_B$ $= \sqrt{2} * \alpha * (\sqrt{2}^{logT} - 1)/(\sqrt{2} - 1)$ (using the equation for sum of a geometric series)
$\Rightarrow$Regret$_B$ $= \sqrt{2} * \alpha * (\sqrt{T} - 1)/(\sqrt{2} - 1)$
$\Rightarrow$Regret$_B$ $= O(\sqrt{2} * \alpha\sqrt{T}/(\sqrt{2} - 1))$
Thus, B has regret at most $\sqrt{2} * \alpha\sqrt{T}/(\sqrt{2} - 1)$.
Q.E.D.

Ex. 4

Part c

To show: $||.||_1^* = ||.||_\infty$

We know, $||y||_1^* = max_{\sum_i |x_i| \leq 1} x^T.\mathbf{y} = \sum_i x_i y_i$

Now, if we select $\mathbf{x}$ such that if the i^{th} element corresponding to the max. magnitude of y_i is +1 (if $y_i > 0$) or -1 ($y_i < 0$), and 0 otherwise, then we will get $||y||_\infty$ on the RHS.

To show: $||.||_\infty^* = ||.||_1$

We know, $||y||_\infty^* = max_{max_i\{|x_i|\} \leq 1} x^T.\mathbf{y}$

Now, if we select $\mathbf{x}$ such that $x_i = sign(y_i)$, we will essentially get on the RHS, $\sum_i |y_i| = ||y||_1$.

To show: $||.||_2^* = ||.||_2$

We know, $||y||_2^* = max_{||x_2|| \leq 1} x^T.\mathbf{y}$

Now, to maximize the dot product, if $\mathbf{x} = \dfrac{\mathbf{y}}{||\mathbf{y}||_2}$, on the RHS we get $||y||_2$.

Part b

To show: $\mathbf{x}.\mathbf{y} \leq ||\mathbf{x}||.||\mathbf{y}||^*$

We know, $||\mathbf{y}||^* = max_{||x|| \leq 1} x^T.\mathbf{y}$

$\Rightarrow ||\mathbf{y}||^* = max_{||x|| \leq 1} x^T.\mathbf{y} \geq \dfrac{x^T.\mathbf{y}}{||\mathbf{x}||}$ (trivial since it is the max)

$\Rightarrow ||\mathbf{x}||||\mathbf{y}||^* \geq x^T.\mathbf{y}$ (multiplying $||\mathbf{x}||$ to above)

This is the generalized Cauchy-Schwartz Identity.

Part a

To show: $||y||_A^* = ||y||_{A^{-1}}^*$, where A is a p.s.d, i.e., $A \succ 0$.

Since $A \succ 0$, $A^{1/2}$ and $A^{-1/2}$ exist and are both p.s.d.

$||y||_A^* = max_{x^T Ax \leq 1} x^T y$

$\Rightarrow ||\text{y}||_A^* = max_x \dfrac{x^T y}{\sqrt{x^T Ax}}$

$\Rightarrow ||\text{y}||_A^* = max_x \dfrac{(A^{-1/2}x)^T y}{\sqrt{x^T A^{-1/2} A A^{-1/2} x}}$ ($A^{-1/2}$ is symmetric, invertible)

$\Rightarrow ||\text{y}||_A^* = max_x \dfrac{x^T A^{-1/2} y}{||x||_2}$

Now, if $x = A^{-1/2} y$,

$\Rightarrow ||\text{y}||_A^* = \dfrac{||A^{-1/2}y||_2^2}{||A^{-1/2}y||_2}$

$\Rightarrow ||\text{y}||_A^* = ||A^{-1/2}y||_2$

$\Rightarrow ||\text{y}||_A^* = \sqrt{(A^{-1/2}y)^T (A^{-1/2}y)}$

$\Rightarrow ||\text{y}||_A^* = \sqrt{y^T A^- 1 y}$

$\Rightarrow ||\text{y}||_A^* = ||y||_{A^{-1}}^*$
Q.E.D

Ex. 2

Part a A/Q, $\nabla^2 f(\mathbf{x}) \succ \alpha \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T$
$\Rightarrow \nabla^2 f(\mathbf{x}) - \alpha \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T \succ 0$
To show : f is $\alpha - expconcave$ over K
Let $g(\mathbf{x}) = e^{-\alpha f(\mathbf{x})}$
Now, g has to be concave. $\Rightarrow \nabla^2 g(\mathbf{x}) \preceq 0$. (To prove)
Now, $g(\mathbf{x}) = e^{-\alpha f(\mathbf{x})}$
$\Rightarrow \nabla g(\mathbf{x}) = \alpha \nabla f(\mathbf{x}) e^{-\alpha f(\mathbf{x})}$
$\Rightarrow \nabla^2 g(\mathbf{x}) = -\alpha \nabla^2 f(\mathbf{x}) e^{-\alpha f(\mathbf{x})} + \alpha^2 \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T e^{-\alpha f(x)}$
$\Rightarrow \nabla^2 g(\mathbf{x}) = -\alpha e^{-\alpha f(\mathbf{x})} (\nabla^2 f(\mathbf{x}) - \alpha \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T)$
$\Rightarrow \nabla^2 g(\mathbf{x}) \preceq 0$ (since $\nabla^2 f(\mathbf{x}) - \alpha \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T \succ 0$)
Q.E.D.

Part b To show : A strongly convex function is also exp-concave.
Let f($\mathbf{x}$) let a strongly convex function.
By definition of strongly convex,
$\nabla^2 f(\mathbf{x}) \succeq \alpha I \succeq \alpha y y^T$
Now, seeing from **Part a**, if $y = \dfrac{\nabla f(\mathbf{x})}{\beta}$, $y^T = \dfrac{\nabla f(\mathbf{x})^T}{\beta}$
$\Rightarrow \nabla^2 f(\mathbf{x}) \succeq \alpha \dfrac{\nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T}{\beta^2}$
From **Part a**, under this scenario, we know that $f(\mathbf{x})$ is $\dfrac{\alpha}{\beta^2}$ - expconcave.
Q.E.D.

To show : An exp-concave function is not always strongly convex.
Let $f(\mathbf{x}) = \dfrac{-log(y^T \mathbf{x})}{\alpha}$
$\Rightarrow$g($\mathbf{x}$) = e$^{-\alpha f(\mathbf{x})} = y^T \mathbf{x}$
$\Rightarrow \nabla g(\mathbf{x})^2 = 0$
Since $\nabla g(\mathbf{x})^2 \preceq 0$, hence g is concave. Therefore, $f(\mathbf{x})$ is an $\alpha - expconcave$ function.
But, for f($\mathbf{x}$) to be strongly convex, $\nabla^2 f(\mathbf{x}) \succeq \alpha I \succeq \alpha y y^T$
Now, $\nabla f(\mathbf{x}) = \dfrac{-y^T x}{\alpha y^T x}$
$\Rightarrow \nabla^2 f(\mathbf{x}) = \dfrac{y^T}{\alpha y^T x^2}$
Now, $\nabla^2 f(\mathbf{x}) \succeq \alpha I$ is not true for this f($\mathbf{x}$).
Therefore, f($\mathbf{x}$) is not strongly convex for $\alpha - expconcave function$ f($\mathbf{x}$).
Q.E.D.