COS511 HW3

Pranjit Kalita

April 1, 2017

Ex. 4
If f is convex and differentiable and $\nabla f(x) = 0$, then f(x) = min f.
True.
Proof:- Since f is convex and differentiable,
f(y) - f(x) $\geq \nabla f(x) * (y - x)$ (from definition)
$\Rightarrow$f(y) - f(x) ≥ 0 ($\nabla f(x) = 0$)
$\Rightarrow$f(y) $\geq f(x)$
$\Rightarrow$f(x) = min f
Q.E.D.

If f, g are convex functions then f + g is a convex function.
True.
Proof:- Let h(x) = f(x) + g(x).
$h(\alpha x_1 + (1-\alpha)x_2) = f(\alpha x_1 + (1-\alpha)x_2) + g(\alpha x_1 + (1-\alpha)x_2)$
We have to show that -
$f(\alpha x_1+(1-\alpha)x_2)+g(\alpha x_1+(1-\alpha)x_2) \leq \alpha(f(x_1)+g(x_1))+(1-\alpha)(f(x_2)+g(x_2))$
Since f and g are convex functions for two points $x_1, x_2 \in dom(f,g)$ and $\alpha \in [0,1]$, we have -

f($\alpha x_1 + (1-\alpha)x_2) \leq \alpha(f(x_1) + (1-\alpha)f(x_2)$ (1)

g($\alpha x_1 + (1-\alpha)x_2) \leq \alpha(g(x_1) + (1-\alpha)g(x_2)$ (2)

Adding Equations 1 and 2 -
$f(\alpha x_1+(1-\alpha)x_2)+g(\alpha x_1+(1-\alpha)x_2) \leq \alpha(f(x_1)+g(x_1))+(1-\alpha)(f(x_2)+g(x_2))$
Q.E.D.

If f is convex and $\alpha \geq 0$ then $\alpha * f$ is a convex function.
True.

$Proof: -Let\ h(x) = \alpha * f(x)$.
Since f is a convex function,
$f(\beta x + (1-\beta)y) \leq \beta f(x) + (1-\beta)f(y)$, for any $0 \leq \beta \leq 1$
Now, if h were convex, for any two points x and y,
$h(\beta x + (1-\beta)y) \leq \beta h(x) + (1-\beta)h(y)$, for any $0 \leq \beta \leq 1$
$h(\beta x + (1-\beta)y) \leq \beta\alpha * f(x) + (1-\beta)\alpha * f(y)$, for any $0 \leq \beta \leq 1$ Now,
$h(\beta x + (1-\beta)y) = \alpha * f(\beta x + (1-\beta)y) \leq \alpha * (\beta f(x) + (1-\beta)f(y))$
$\Rightarrow$h$(\beta x + (1-\beta)y) \leq \alpha * \beta * f(x) + \alpha * (1-\beta) * f(y)$
Thus, h is convex.
Q.E.D.

If F is a set of convex functions and $F_m = \max_{f \in F}(f)$ **is a convex function**.
True.
Proof:- Let us prove for two functions f and g, i.e., let $h = maxf, g$, where f and g are both convex functions. We can easily extend to multiple functions and come to the conclusion.
Since f is convex,

f$(\beta x + (1-\beta)y) \leq \beta f(x) + (1-\beta)f(y)$, for any $0 \leq \beta \leq 1(3)$

Since g is convex,

g$(\beta x + (1-\beta)y) \leq \beta g(x) + (1-\beta)g(y)$, for any $0 \leq \beta \leq 1(4)$

Now if h is convex, we will require,
$h(\beta x + (1-\beta)y) \leq \beta h(x) + (1-\beta)h(y)$, for any $0 \leq \beta \leq 1$
We know, since h = maxf,g,

f(x) $\leq h(x)(5)$

f(y) $\leq h(y)(6)$

Similarly, for g(x) and g(y).
Thus, combining equations 3, 4, 5, 6 -
$f(\beta x + (1-\beta)y) \leq \beta h(x) + (1-\beta)h(y)$, for any $0 \leq \beta \leq 1$
$g(\beta x + (1-\beta)y) \leq \beta h(x) + (1-\beta)h(y)$, for any $0 \leq \beta \leq 1$
Since h = maxf,g, thus we get -
$h(\beta x + (1-\beta)y) \leq \beta h(x) + (1-\beta)h(y)$, for any $0 \leq \beta \leq 1$
Therefore, h is convex.
Applying it to multiple functions instead of 2 functions the result can be trivially extended to prove for any number of functions.
Q.E.D.

If K and G are convex sets then $K + G = \{u + v : u \in K, v \in G\}$ is a convex set.
True.
Proof: Let x$_k, y_k \in K$.
$\Rightarrow \alpha x_k + (1-\alpha)y_k \in K, 0 \leq \alpha \leq 1$
$Let x_g, y_g \in G.$
$\Rightarrow \alpha x_g + (1-\alpha)y_g \in G, 0 \leq \alpha \leq 1$
Now, (x$_k + x_g), (y_k + y_g) \in K + G$.
We have to prove that
$\alpha(x_k + x_g) + (1-\alpha)(y_k + y_g) \in K + G$
$\alpha(x_k + x_g) + (1-\alpha)(y_k + y_g)$
$= \alpha x_k + (1-\alpha)y_k + \alpha x_g + (1-\alpha)y_g$
$\Rightarrow \alpha x_k + (1-\alpha)y_k + \alpha x_g + (1-\alpha)y_g \in K + G$ $(\alpha x_k + (1-\alpha)y_k \in K, \alpha x_g + (1-\alpha)y_g \in G)$
Thus, K + G is convex set.
Q.E.D.

If for every the set $A_\alpha = \{x : f(x) < \alpha\}$ is convex, the f is convex.
False.
Proof: Let u, v be such that $f(u) < \alpha$, $f(v) < \alpha$.
Therefore, $u, v \in A_\alpha$
Since A_α is convex set, $\lambda u + (1-\lambda)v \in A_\alpha$, for $0 \leq \lambda \leq 1$.
$\Rightarrow f(\lambda u + (1-\lambda)v) < \alpha$
$\Rightarrow$f$(\lambda u + (1-\lambda)v) < \lambda\alpha + (1-\lambda)\alpha$
Now, to get $f(\lambda u + (1-\lambda)v) < \lambda f(u) + (1-\lambda)f(v)$
$\alpha = f(u)$, $\alpha = f(v)$, which is not possible according to the definition of the set A_α. Moreover, u and v are two different points.
Thus, we arrive at a contradiction, hence this is not true.
Q.E.D.

If f, g are convex functions then $f \circ g$ is a convex function.
True (only if f is a nondecreasing function).
Proof: Since f is a convex function, for $0 \leq \lambda \leq 1$,
$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$

Similarly, for g,
$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$
$f \circ g(\lambda x + (1-\lambda)y) = f(g(\lambda x + (1-\lambda)y))$
$\Rightarrow$f $\circ g(\lambda x + (1-\lambda)y) \leq f(\lambda g(x) + (1-\lambda)g(y))$
(since g is convex)
At this stage, if f were not nondecreasing, then we cannot proceed to the next step, and hence we will have proven that $f \circ g$ is not convex under that condition.
If f were nondecreasing,
$\Rightarrow$f $\circ g(\lambda x + (1-\lambda)y) \leq \lambda f(g(x)) + (1-\lambda)f(g(y))$
$\Rightarrow$f $\circ g(\lambda x + (1-\lambda)y) \leq \lambda f \circ g(x) + (1-\lambda)f \circ g(y)$
Thus, $f \circ g$ is convex under these circumstances.
Q.E.D.

Ex. 1

The given statement does not hold under all circumstances, and in order to provide a counter example, we will use a trivial instance of a group of hypothesis classes **H** without any assumptions (i.e, in theory they could be non-uniformly convegent).

Now, let us add to **H** a hypothesis h' such that $\sum_{i=1}^{m} l(h, z) = inf_{h* \in \mathbf{H}} \sum_{i=1}^{m} l(h*, z)$

Now, with an ERM algorithm, it will always return h'.

Hence, the given statement does not hold true for non-binary classification problems, in trivial learning examples.

Ex. 3

a

$R_S(cF) = \mathbf{E}_\sigma[sup_{f\in F}|(1/m) * \sum_{i=1}^{m} \sigma_i c * f(z^{(i)})|]$
$\Rightarrow \mathrm{R}_S(cF) = \mathbf{E}_\sigma[sup_{f\in F}|(1/m) * c * \sum_{i=1}^{m} \sigma_i * f(z^{(i)})|]$
$\Rightarrow \mathrm{R}_S(cF) = \mathbf{E}_\sigma[sup_{f\in F}|c| * |(1/m) * \sum_{i=1}^{m} \sigma_i * f(z^{(i)})|]$
$\Rightarrow \mathrm{R}_S(cF) = |c| * \mathbf{E}_\sigma[sup_{f\in F}|(1/m) * \sum_{i=1}^{m} \sigma_i * f(z^{(i)})|]$
$\Rightarrow \mathrm{R}_S(cF) = |c| * R_S(F)$
Q.E.D.

b

$R_S(convF) = \mathbf{E}_\sigma[sup_{f'\in convF}|(1/m) * \sum_{i=1}^{m} \sigma_i f'(z^{(i)})|]$
$\Rightarrow \mathrm{R}_S(convF) = \mathbf{E}_\sigma[sup_{\sum \lambda_j f_j \in convF}|(1/m) * \sum_{i=1}^{m} \sigma_i \sum \lambda_j f_j(z^{(i)})|]$
$\Rightarrow \mathrm{R}_S(convF) = \mathbf{E}_\sigma[sup_{f\in F}|(1/m) * \sum \lambda_j \sum_{i=1}^{m} \sigma_i f(z^{(i)})|]$
$\Rightarrow \mathrm{R}_S(convF) = \sum \lambda_j * \mathbf{E}_\sigma[sup_{f\in F}|(1/m) * \sum_{i=1}^{m} \sigma_i f(z^{(i)})|]$
$\Rightarrow \mathrm{R}_S(convF) = \mathbf{E}_\sigma[sup_{f\in F}|(1/m) * \sum_{i=1}^{m} \sigma_i f(z^{(i)})|]$ (since $\sum \lambda_j = 1$)
$\Rightarrow \mathrm{R}_S(convF) = R_s(F)$
Q.E.D.

Ex. 6
To prove: $l(\lambda u + (1-\lambda)w, (x,y)) \leq \lambda l(u,(x,y) + (1-\lambda)l(w,(x,y))$
$l(\lambda u + (1-\lambda)w, (x,y)) = min\{\sum_{i=1}^{2^d} \alpha_i l(v_i,(x,y), \sum \alpha_i = 1, \alpha_i \geq 0, \lambda u + (1-\lambda)w = \sum_{i=1}^{2^d} \alpha_i v_i\}$
$l(\lambda u + (1-\lambda)w, (x,y)) = \sum_{i=1}^{i=2^d} \alpha_i^* l(v_i,(x,y))$ (let)
$l(u,(x,y) = \sum_{i=1}^{i=2^d} \alpha_i^u l(v_i,(x,y))$ (let)
$l(u,(x,y) = \sum_{i=1}^{i=2^d} \alpha_i^w l(v_i,(x,y))$ (let)
$\Rightarrow \lambda l(u,(x,y) + (1-\lambda)l(w,(x,y) = \lambda * \sum_{i=1}^{i=2^d} \alpha_i^u l(v_i,(x,y)) + (1-\lambda) * \sum_{i=1}^{i=2^d} \alpha_i^w l(v_i,(x,y))$
$\Rightarrow \lambda l(u,(x,y) + (1-\lambda)l(w,(x,y) = \sum_{i=1}^{i=2^d} \beta_i l(v_i,(x,y))$ (where $\beta_i = \lambda * \alpha_i^u + (1-\lambda) * \alpha_i^w$)
So, essentially, we have to prove -
$\sum_{i=1}^{i=2^d} \alpha_i^* l(v_i,(x,y)) \leq \sum_{i=1}^{i=2^d} \beta_i l(v_i,(x,y))$
Thus, if β satisfies $\sum_{i=1}^{2^d} \beta_i = 1$, $\beta_i \geq 0$, $\mathbf{w} = \sum_{i=1}^{2^d} \beta_i v_i$ (where $\mathbf{w}$ is any d-dimensional vector $\in [0,1]^d$), we will have proven the proof.
$\sum_{i=1}^{2^d} \beta_i = \lambda \sum_{i=1}^{2^d} \alpha_i^u + (1-\lambda) \sum_{i=1}^{2^d} \alpha_i^w$
$\Rightarrow \sum_{i=1}^{2^d} \beta_i = \lambda + (1-\lambda)$ ($\sum_{i=1}^{2^d} \alpha_i^u = 1$, $\sum_{i=1}^{2^d} \alpha_i^w = 1$)
$\Rightarrow \sum_{i=1}^{2^d} \beta_i = 1$ (Property 1 proved)
$\beta_i = \lambda \alpha_i^u + (1-\lambda)\alpha_i^w \geq 0$ ($0 \leq \lambda \leq 1$, $\alpha_i^u \geq 0$, $\alpha_i^w \geq 0$)
Property 2 proved
$\mathbf{w} = \sum_{i=1}^{2^d} \beta_i v_i$
$\Rightarrow \mathbf{w} \in [0,1]^d$ (since each dimension of v$_i is 0 or 1, and 0 \leq \beta_i \leq 1$, $\sum_{i=1}^{2^d} \beta_i = 1$, therefore in $\mathbf{w}$ each dimension lies in the set [0,1]).
Therefore, we see that $\sum_{i=1}^{2^d} \beta_i v_i$ leads to a d-dimensional vector in $[0,1]^d$.
Therefore, Property 3 proved.
Thus, β is essentially similar to α^* in the context of l function.
Thus, $\sum_{i=1}^{i=2^d} \alpha_i^* l(v_i,(x,y)) \leq \sum_{i=1}^{i=2^d} \beta_i l(v_i,(x,y))$
$\Rightarrow$l$(\lambda u + (1-\lambda)w, (x,y)) \leq \lambda l(u,(x,y) + (1-\lambda)l(w,(x,y))$
Q.E.D.