COS511 HW4

Pranjit Kalita

April 22, 2017

Ex. 1

2

Let us assume that there exists y_1 and y_2 which are both projections of $\mathbf{x}$ onto a convex set K.
Also, y1 $\neq y2$
y_1 =argmin$_{\mathbf{y}\in K}||\mathbf{x}-\mathbf{y}||$
y_1 =argmin$_{\mathbf{y}\in K}||\mathbf{x}-\mathbf{y}||$
We will try to arrive at a contradiction with our premise, and by using the following algebraic identity -
$1/2*(||u+w||^2) = ||u||^2 + ||v||^2 - 1/2*||u-v||^2$
Let, $u = \mathbf{x} - y_1$, $w == \mathbf{x} - y_2$
$\Rightarrow$1/2 * (||2$\mathbf{x}$ - (y$_1 + y_2$)||)^2 $= ||\mathbf{x}-y_1||^2 + ||\mathbf{x}-y_|||^2 - 1/2*||y_2-y_1||^2$
$\Rightarrow$1/4 * (||2$\mathbf{x}$ - (y$_1 + y_2$)||)^2 $= 1/2*||\mathbf{x}-y_1||^2 + 1/2*||\mathbf{x}-y_2||^2 - 1/4*||y_2-y_1||^2$
$\Rightarrow$||((2$\mathbf{x}$ - (y$_1 + y_2$))/2||^2 $< 1/2*||\mathbf{x}-y_1||^2 + 1/2*||\mathbf{x}-y_2||^2$
$\Rightarrow$(||$\mathbf{x}$ - (y$_1 + y_2$)/2||)^2 $< 1/2*||\mathbf{x}-y_1||^2 + 1/2*||\mathbf{x}-y_2||^2$
But, since y_1, y_2 =argmin$_{\mathbf{y}\in K}||\mathbf{x}-\mathbf{y}||$
$||\mathbf{x}-y_1|| = ||\mathbf{x}-y_2||$
$\Rightarrow$||$\mathbf{x}$ - (y$_1 + y_2$)/2||^2 $< ||\mathbf{x}-y_1||^2$
Now, since $y_1, y_2 \in K$, and K is a convex set, $0.5y_1 + 0.5y_2 \in K$ ($\lambda = 0.5$ in this case acc. to definition of convex set)
But, according to our premise, $y_1$andy_2 were projections, but we found $0.5y_1 + 0.5y_2 \in K$ which is not equal to y$_1$ory_2,whose value
||x - (0.5y$_1 + 0.5y_2$)|| $< ||x - y_1||$, thereby meaning that y_1 and y_2 both being the minimum value (thereby the projections) is being contradicted.
Therefore we arrive at a contradiction with the premise that both y_1 and y_2 were projections at the same time.
Thus, there can be only one projection of $\mathbf{x}$ onto a convex set K.
Q.E.D

1

Given a convex set K, and a function $\pi_K(\mathbf{y}) = argmin_{\mathbf{x}\in K}||\mathbf{x}-\mathbf{y}||$
To prove : $\forall \mathbf{z} \in K$,||$\mathbf{z}$ - $\pi_K(\mathbf{y})$|| $< ||\mathbf{z}-\mathbf{y}||$
Proof : This is a fairly trivial proof. For our proof, we will make use

of the definition of $\pi_K(\mathbf{y}) = argmin_{\mathbf{x}\in K}||\mathbf{x} - \mathbf{y}||$.
$||$**z** - **x**$||, where \mathbf{x} \in K$,$\mathbf{x} = \pi_K(\mathbf{y})$ is one side of the triangle formed by the vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, where $\mathbf{z}, \mathbf{x} \in K$, $\mathbf{y} \notin K$.
We know that ||y - x|| will be the shortest side if we join the sides of the triangle formed by the vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, where $\mathbf{z}, \mathbf{x} \in K$, $\mathbf{y} \notin K$.
.Now, if we were to visualize the above triangle, the angle joining $||\mathbf{z} - \mathbf{x}||$ and $||\mathbf{y} - \mathbf{x}||$ is obtuse.
By opposing angle property, the side opposite that obtuse angle $||\mathbf{z}-\mathbf{y}||$ will be the largest in the traingle.
Therefore, $||\mathbf{z} - \pi_K(\mathbf{y})|| < ||\mathbf{z} - \mathbf{y}||$.
Q.E.D

Ex. 3 - Consulted Book 4, Example 16.1 (similar question and reasoning)
To prove : The following homogeneous polynomial kernel is indeed a kernel - k($\mathbf{x}^{(1)}$,$\mathbf{x}^{(2)}$) = $(\mathbf{x}^{(1)},\mathbf{x}^{(2)})^d$
Proof : We have to show that
k($\mathbf{x}^{(1)}$,$\mathbf{x}^{(2)}$) = <$\phi(\mathbf{x}^{(1)})$, $\phi(\mathbf{x}^{(2)}$>,where ϕ is a mapping from the original sample space to a higher dimensional space.
We have -
k($\mathbf{x}^{(1)}$,$\mathbf{x}^{(2)}$) = ($\mathbf{x}^{(1)}$,$\mathbf{x}^{(2)}$).....($\mathbf{x}^{(1)}$,$\mathbf{x}^{(2)}$)
$\Rightarrow$**k($\mathbf{x}^{(1)}$,$\mathbf{x}^{(2)}$) =** $\sum_{j=0}^{n}(x_j^{(1)}, x_j^{(2)})..... \sum_{j=0}^{n}(x_j^{(1)}, x_j^{(2)})$**, where** $\mathbf{x}_0^{(1)} = x_0^{(2)} = 0$
$\Rightarrow$**k($\mathbf{x}^{(1)}$,$\mathbf{x}^{(2)}$) =** $\sum_{J\in\{0,1,2,..,n\}^k} \prod_{i=1}^{k} x_{J_i}^{(1)} x_{J_i}^{(2)}$
$\Rightarrow$**k($\mathbf{x}^{(1)}$,$\mathbf{x}^{(2)}$) =** $\sum_{J\in\{0,1,2,..,n\}^k} \prod_{i=1}^{k} x_{J_i}^{(1)} \prod_{i=1}^{k} x_{J_i}^{(2)}$
$\Rightarrow$**k($\mathbf{x}^{(1)}$,$\mathbf{x}^{(2)}$) = <$\phi(\mathbf{x}^{(1)})$, $\phi(\mathbf{x}^{(2)})$>**
where J $\in \{0,1,...,n\}^k$**,** $\phi : R^n -> R^{(n+1)^k}, \phi(\mathbf{x}_j^{(1)}) = \prod_{i=1}^{k} x_{J_i}^{(1)}$**is each of the elements of the (k+1) dimensional ϕ vector.**
ϕ contains monomials up to degree k, and a halfspace over the range of ϕ yields a polynomial predictor of degree k over the sample space, which can then be used as an embedded linear halfspace predictor.
Q.E.D.

Ex. 2 - Proof similar to textbook 4 (lemmas and theorems referenced are from Book 4).
From Equation 14.2, we have
$f(\bar{w}) - f(w^*) = 1/T * \sum_{t=1}^{T}(f(w^{(t)}) - f(w^*))$, **where** $\mathbf{w}^* = agrmin_{w\in K} f(w)$
$\Rightarrow \mathbf{E}_{v1:T}(\mathbf{f}(\bar{w}) - \mathbf{f}(\mathbf{w}^*)) = \mathbf{E}_{v1:T}(1/T * \sum_{t=1}^{T}(f(w^{(t)}) - f(w^*)))$
Acc. to Lemma 14.1
$1/T * \sum_{t=1}^{T} < w^{(t)} - w^*, v_t > \leq DG/\sqrt{T}$, **where** $\eta = D/G\sqrt{T}$
$\Rightarrow \mathbf{E}_{v_{1:T}}(1/T * \sum_{t=1}^{T} < w^{(t)} - w^*, v_t >) \leq E(DG/\sqrt{T})$
$\Rightarrow \mathbf{E}_{v_{1:T}}(1/T * \sum_{t=1}^{T} < w^{(t)} - w^*, v_t >) \leq DG/\sqrt{T}$
$\Rightarrow$ **1/T * $\mathbf{E}_{v_{1:T}}(\sum_{t=1}^{T} < w^{(t)} - w^*, v_t >) \leq DG/\sqrt{T}$ (by linearity of expectation)**
Thus, we have to show that -
$E_{v_{1:T}}(1/T * \sum_{t=1}^{T}(f(w^{(t)}) - f(w^*))) \leq 1/T * E_{v_{1:T}}(\sum_{t=1}^{T} < w^{(t)} - w^*, v_t >)$
By applying law of total expectations: $E_\alpha[g(\alpha)] = E_\beta E_\alpha[g(\alpha)|\beta]$
$E_{v_{1:T}}(\sum_{t=1}^{T} < w^{(t)} - w^*, v_t >) = E_{v_{1:t-1}} E_{v_{1:t}}[< w^{(t)} - w^*, v_t > |v_{1:t-1}]$
Now, $E_{v_{1:T}}(\sum_{t=1}^{T} < w^{(t)} - w^*, v_t >) = E_{v_{1:t-1}} < w^{(t)} - w^*, E_{v_t}[v_t|v_{1:t-1}]$
Since the value of $w^{(t)}$ **depends on** $v_{1:t-1}$ **and in SGD, at each step** $E_{v_t}[v_t|w_t] \in \partial f(w^{(t)})$,
We have
$E_{v_t}[v_t|v_{1:t-1}] \in \partial f(w^{(t)})$
Now, using Equation 14.3 of textbook and using convexity of f,
$E_{v_{1:t-1}} < w^{(t)} - w^*, E_{v_t}[v_t|v_{1:t-1}] > \geq E_{v_{1:t-1}}[f(w^{(t)}) - f(w^*)]$
$\Rightarrow \mathbf{E}_{v_{1:T}} < w^{(t)} - w^*, v_t > \geq E_{v_{1:T}}[f(w^{(t)}) - f(w^*)]$
Which is the proof we needed.
Therefore, $E[f(w^{(t)}) - f(w^*)] \leq DG/\sqrt{T}$
$\Rightarrow \mathbf{f}(\mathbf{w}^{(t)}) \leq f(w^*) + DG/\sqrt{T}$
Q.E.D

Ex. 4

We will begin this problem by observing that Δ has the same number of μ functions comprising it as F (since it is a class of distributions over [r]).

Next, we see that $\mu(i)$ and $f(i)$ have 1-1 correspondence, since μ is used to find distributions over each feature function f.

Now, let us find $|F| = |\Delta|$.

Since F is the class of all monomials of degree d over $\mathbf{X} = \mathbf{R}^n$.

$Thus, any$**f** $\in F$, $X^d = x_1^{d1} x_2^{d2} .... x_n^{dn}$, **where** $\sum_{i=1}^{n} d_i = d$

Thus, in order to find the number of functions in F, we can think of the above as placing d integer's partitions into (n-1) spots, which zeroes allows.

We can see that the actual value of the total no of functions $(r) = C((n-1+d), n-1)$

Thus, $r = |F| = |\Delta| = C((n-1+d), n-1)$

Further simplifying, $|\Delta| = (n+d-1)(n+d-2)(n+d-3_{.}...(n+d-d)/d! \leq n^d$

$\Rightarrow |\Delta| \leq n^d$

Δ **is infinite. However, the convex hull of** $\Delta = \mu$ **is finite (acc. to definition of convex hull, given the problem's specifications** $\sum_{i=1}^{r} \mu(i) = 1$, $\mu(i) > 0$, $\mu(x) = \sum \mu(i) f(i)$, $f \in F$**).**

Therefore, now we can use Corollary 12.2 from lecture notes:

$R_m(\Delta) \leq \sqrt{2log(n^d)/m}$

$\Rightarrow \mathbf{R}_m(\Delta) \leq \sqrt{2dlog(n)/m}$

Now, l is L-Lipschitz loss function acc. to the problem, therefore using the Lemma 12.4 in notes, we have

$R_m(l \circ \Delta) \leq L * R_m(\Delta)$

$\Rightarrow \mathbf{R}_m(l \circ \Delta) \leq L * \sqrt{2dlog(n)/m}$

$\Rightarrow \mathbf{R}_m(l \circ \Delta) = O(L * \sqrt{2dlog(n)/m})$

Thus, sample complexity is of the required order.

Q.E.D.