Note Title 19-11-2020

COL774
Minor solutions

Q.1. (a) The general form of the degree k classifier can be written as:- (assume binary class labels):

$$\sum_{\substack{t_1 \ldots t_n \\ \text{s.t. } t_1+t_2+\ldots t_n \leq k}} \theta_{t_1 \ldots t_n} \prod_{j=1}^{n} x_j^{t_j} = 0$$

Note:- this is the most general form of a polynomial with degree k (each term (monomial) in the expression is degree K or less and all such possible terms are constructed

(b) $\phi(x) = \sum_{k=0}^{k} a_k x^k \equiv \sum_{k=0}^{k} a_k (x_0, x_1, \ldots, x_n)^k$

which will generate all the monomial terms of degree k or less features.

It is given that $\phi(x)^T \phi(z) = K(x, z)$

$= (x^T z + c)^d$

missing in question but not really req. for solution

Now, let us look at the ~~w~~ parameter update equation:-

[Note:- entire algorithm is now run in the transformed feature space i.e. represented by $\phi(x)$]

Training Data:- $\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left[y^{(i)} - h_\theta(\phi(x^{(i)}))\right]^2$

in transformed space

$\theta^{(0)} \leftarrow \vec{0}$ [initialization to 0 vector]

$t = 0$

do {

$\theta^{(t+1)} \leftarrow \theta^{(t)} - \eta/_m \sum_{i=1}^{m}\left[y^{(i)} - h_{\theta}(\phi(x^{(i)}))\right]\phi(x^{(i)})$

$t \leftarrow t+1$

} (until convergence);

[replaced occurrence of $x^{(i)}$ by $\phi(x^{(i)})$]

Now:- $$h_\theta(\phi(x^{(i)})) = \frac{1}{1 + e^{-\theta^T \cdot \phi(x^{(i)})}}$$

We will assume that θ can be represented as coefficients $\alpha_1 \ldots \alpha_m$ of data points $x^{(1)} \ldots x^{(m)}$. In particular:-

$$\theta = \sum_{i=1}^{m} \alpha_i \phi(x^{(i)})$$

At time t, $$\theta^{(t)} = \sum_{i=1}^{m} \alpha_i^{t} \phi(x^{(i)})$$

We will show that this assumption holds inductively

At $t=0$, this assumption is true. Since,
at $t=0$, $\theta^{(t)}=0 \Rightarrow \alpha_i^t = 0 \ \forall i$.

Next assume it holds for ~~the~~ step t.
Then let us see if we can prove for
step $t+1$.

$$\theta^{(t+1)} \leftarrow \theta^{(t)} - \eta \cdot 1/m \sum_{i=1}^{m} \left[y^{(i)} - h_{\theta^{(t)}}(\phi(x^{(i)})) \right] x^{(i)}$$

Then: $$h_{\theta^{(t)}}(x^{(i)}) = \frac{1}{1 + e^{-\theta^{(t)T} \cdot \phi(x^{(i)})}}$$

Now, by inductive assumption, we have $\theta^{(t)} = \sum_{i=1}^{m} \alpha_i^t \phi(x^{(i)})$

$$\Rightarrow \quad \theta^{(t)T} \cdot \phi(x^{(i)}) = \left[\sum_{l=1}^{m} \alpha_l^t \, \phi(x^{(l)})\right]^T \cdot \phi(x^{(i)})$$

→ change of variable from i to l

Computation only in terms of $\phi(x^{(l)})^T \phi(x^{(i)})$ terms.

No other $\phi(x^{(i)})$ involvement in the expression

$$= \sum_{l=1}^{m} \alpha_l^t \left[\phi(x^{(l)})^T \phi(x^{(i)})\right]$$

$$= \sum_{l=1}^{m} \alpha_l^t \, K(x^{(l)}, x^{(i)}) \equiv \beta_i^t$$

$$\Rightarrow \quad \theta^{(t+1)} \leftarrow \theta^{(t)} - \eta \cdot 1/m \sum_{i=1}^{m} \left[y^{(i)} - \frac{1}{1+e^{-\beta_i^t}}\right] \phi(x^{(i)})$$

γ_i^t

$$\Rightarrow \quad \theta^{(t+1)} \leftarrow \theta^{(t)} - \eta \cdot \frac{1}{m} \sum_{i=1}^{m} \gamma_i^t \, \phi(x^{(i)})$$

$$\Rightarrow \theta^{(t+1)} \leftarrow \theta^{(t)} - \sum_{i=1}^{m} \gamma_i'^t \cdot \phi(x^{(i)})$$

$$\left[\gamma_i'^t = \eta/m \, \gamma_i^t\right]$$

$\Rightarrow$ Using inductive assumption,

$$\theta^{(t)} = \sum_{i=1}^{m} \alpha_i^t \, \phi(x^{(i)})$$

$$\Rightarrow \quad \theta^{(t+1)} = \sum_{i=1}^{m} \alpha_i^t \, \phi(x^{(i)}) - \sum_{i=1}^{m} \gamma_i'^t \, \phi(x^{(i)})$$

$$\Rightarrow \theta^{(t+1)} = \sum_{i=1}^{m} \underbrace{(\alpha_i^t - \gamma'^t_i)}_{\text{Define } \alpha_i^{t+1}} \phi(x^{(i)})$$

$$\Rightarrow \theta^{(t+1)} = \sum_{i=1}^{m} \alpha_i^{(t+1)} \phi(x^{(i)})$$

$\Rightarrow$ Hence, inductive assumption is satisfied.

Further, the entire computation can be done in form of $[\phi(x)^T \phi(z)]$ the expressions only. since we did not use $\phi(x^{(i)})$ in any other form on above computation.

(c) Computational complexity:

$\phi(x)$ will have exponential (in k) number of terms as features [All possible monomials in $x_1 \ldots x_n$ of degree $\le k$]

Therefore, ~~we~~ if we directly represent $\phi(x)$ it will consume exponential in k time/memory.

But if we only need to compute

$$K(x,z) = \phi(x)^T\phi(z) = (x^Tz+c)^d$$

this takes only $O(n+d)$ time. (& memory in $O(n)$)