COMS 4771 Machine Learning (Spring 2020) Problem Set #1

Angela Wang - `aw3062@columbia.edu`

02/21/2020

Problem 3

(i)

(ii)

(iii)

Problem 4

(i)

Proof. Since $M^T = (A^T A)^T = A^T A = M$, M is symmetric. For any column vector $z \in \mathbb{R}^d$, $z^T M z = z^T A^T A z = (Az)^T (Az) = (Az) \cdot (Az) \geq 0$, so M is positive semi-definite. $\qquad\square$

(ii)

Proof. Base case: if $N = 1$, then by definition,

$$\beta^{(1)} = \beta^{(0)} + \eta A^T (b - A\beta^{(0)})$$
$$= \eta v - \eta M \beta^{(0)}$$
$$= \eta v$$

Assuming the statement $\beta^{(N)} = \eta \sum_{k=0}^{N-1} (I - \eta M)^k v$ is true for $N = n$, then for $N = n + 1$ we have

$$\beta^{(n+1)} = (I - \eta M)\beta^{(n)} + \eta v \qquad \text{(definition)}$$
$$= (I - \eta M)\eta \sum_{k=0}^{n-1} (I - \eta M)^k v + \eta v \qquad \text{(induction hypothesis)}$$
$$= \eta \sum_{k=0}^{n-1} (I - \eta M)^{(k+1)} v + \eta (I - \eta M)^0 v$$
$$= \eta \sum_{k=1}^{n} (I - \eta M)^{(k)} v + \eta (I - \eta M)^0 v$$
$$= \eta \sum_{k=0}^{n} (I - \eta M)^{(k)} v$$

$\qquad\square$

(iii)

Since M is a real symmetric matrix, it has a decomposition $M = PDP^{-1}$, where P is a matrix composed of orthogonal eigenvectors corresponding to distinct eigenvalues, and

$D := \mathrm{diag}(\lambda_1, \ldots, \lambda_d).$

$$\eta \sum_{k=0}^{N-1} (I - \eta M)^k$$

$$= \eta \sum_{k=0}^{N-1} \sum_{i=0}^{k} I^{k-i}(-\eta M)^i$$

$$= \sum_{k=0}^{N-1} \sum_{i=0}^{k} (-1)^i \eta^{i+1} M^i$$

$$= \sum_{k=0}^{N-1} \sum_{i=0}^{k} (-1)^i \eta^{i+1} P D^i P^{-1}$$

$$= P \left(\sum_{k=0}^{N-1} \sum_{i=0}^{k} (-1)^i \eta^{i+1} \mathrm{diag}(\lambda_1^i, \ldots, \lambda_d^i) \right) P^{-1}$$

$$= P \, \mathrm{diag} \left(\sum_{k=0}^{N-1} \sum_{i=0}^{k} (-1)^i \eta^{i+1} \lambda_1^i, \ldots, \sum_{k=0}^{N-1} \sum_{i=0}^{k} (-1)^i \eta^{i+1} \lambda_d^i \right) P^{-1}$$

$$= P \, \mathrm{diag} \left(\sum_{i=0}^{N-1} (N-i)(-1)^i \eta^{i+1} \lambda_1^i \ldots, \sum_{i=0}^{N-1} (N-i)(-1)^i \eta^{i+1} \lambda_d^i \right) P^{-1}$$

Since I, M are symmetric, $\eta \sum_{k=0}^{N-1} (I - \eta M)^k$ is symmetric, and its eiganvalues $\lambda_1', \ldots, \lambda_d'$ are given by

$$\lambda_j' = \sum_{i=0}^{N-1} (N-i)(-1)^i \eta^{i+1} \lambda_j^i$$

for any $j = 1, \ldots, d$.

(iv)

1 Problem 5

Part 0

Ignoring the sensitive attribute does not guarantee fairness because there may be other attributes with correlation to that sensitive attributes. For instance, assume that the sensitive attribute is income level, but you also have attributes like number of cars owned, average spending per month, etc. Simply removing income level does not prevent the model from maintaining a bias since it will still use number of cars owned, average spending per month, etc. to classify the data. As a result, the model will likely fail to maintain fairness across the sensitive distribution.

Part 0

(i)

$$P[Y = 1] = P_a[Y = 1]$$

$$P_a[Y = 1] = P[Y = 1|a = 0] = P[Y = 0|a = 1]$$

$$P[Y = 1|a = 0] = P[Y = 0|a = 1]$$

$$P_0[Y = 1] = P_1[Y = 1]$$

1.0.1 (ii))

Generalizing:

$$P_a[Y = 1] = P_{a'}[Y = 1] \forall a, a' \in A$$
$$P[y = 1] = P_a[y = 1] \forall a \in A$$