Additional pertinent questions from AY 21/22 CA 2

Answers follow on subsequent page. For most effectiveness, do the questions first without referencing the model answers.

[Questions 1–3] (MCQ; 2 marks each) Mark (a) for true and (b) for false for each of the following statements on **Decision Trees**.

Let's examine decision tree learning as taught in lecture, with categorical inputs X and output Y. Here we assume we do not employ pruning.

1. The depth of the tree cannot exceed $n + 1$.
2. If $IG(Y|X_i) = 0$, then X_i will not be used in the decision tree.
3. Suppose one of the attributes has a unique value in each instance. Then the decision tree must have depth 0 or 1.

4. (MCQ; 2 marks) We stated that validation generally produces an optimistic estimate of L_{test}. Why?

(a) Because we choose the model based on their performance.
(b) Because possibly many parameters are tested in the validation process.
(c) Because possibly many values of parameters are tested in the validation process.

[Questions 5–6] In lecture, we have seen a soft order constraint of the form $\boldsymbol{\theta}^\top\boldsymbol{\theta} < C$. For some problems, there exists a more general soft order constraint of the form $\boldsymbol{\theta}^\top\boldsymbol{M}^\top\boldsymbol{M}\boldsymbol{\theta} < C$ which captures the relationship among the θ_i. Let us fix the constant C and $\boldsymbol{\theta} = [\theta_1,...,\theta_n]^\top{}_{1\times n}$

5. (MCQ; 4 marks) Define $\boldsymbol{M}$ such that we obtain the constraint $\sum \theta_i^2 < 4C$.

(a) $[1,1,...,1]^\top{}_{n\times 1}/4$.
(b) $\boldsymbol{I}_{n\times n}/4$.
(c) $[1,1,...,1]_{1\times n}/4$.
(d) $\boldsymbol{I}_{n\times n}/8$.
(e) None of the above.

6. (MCQ; 4 marks) Define $\boldsymbol{M}$ such that we obtain the constraint $(\sum \theta_i)^2 < C$.

(a) $[1,1,...,1]^\top{}_{n\times 1}$.
(b) $\boldsymbol{I}_{n\times n}$.
(c) $[1,1,...,1]_{1\times n}$.
(d) $\boldsymbol{I}_{n\times n}/4$.
(e) None of the above.

1. The depth of the tree cannot exceed $n + 1$.

   Correct answer: (a) True because the attributes are categorical and can each be split only once.

2. If $IG(Y|X_i) = 0$, then X_i will not be used in the decision tree.

   Correct answer: (b) The attribute may have non-zero IG when used further down in the decision tree in subproblems, after splitting on other attributes.

3. Suppose one of the attributes has a unique value in each instance. Then the decision tree must have depth 0 or 1.

   Correct answer: (a) True because that attribute will have perfect information gain. If an attribute has perfect information gain it must split the records into "pure" buckets which can be split no more.

4. (MCQ; 2 marks) We stated that validation generally produces an optimistic estimate of L_{test}. Why?

(a) Because we choose the model based on their performance.
(b) Because possibly many parameters are tested in the validation process.
(c) Because possibly many values of parameters are tested in the validation process.

Correct answer: (a)

[Questions 5–6] In lecture, we have seen a soft order constraint of the form $\boldsymbol{\theta}^\top\boldsymbol{\theta} < C$. For some problems, there exists a more general soft order constraint of the form $\boldsymbol{\theta}^\top\boldsymbol{M}^\top\boldsymbol{M}\boldsymbol{\theta} < C$ which captures the relationship among the θ_i. Let us fix the constant C and $\boldsymbol{\theta} = [\theta_1,...,\theta_n]^\top{}_{1\times n}$

4. (MCQ; 4 marks) Define $\boldsymbol{M}$ such that we obtain the constraint $\sum \theta_i^2 < 4C$.

(a) $[1,1,...,1]^\top{}_{n\times 1}/4$.
(b) $\boldsymbol{I}_{n\times n}/4$.
(c) $[1,1,...,1]_{1\times n}/4$.
(d) $\boldsymbol{I}_{n\times n}/8$.
(e) None of the above.

Correct answers: (e)

Explanation: $M = \boldsymbol{I}_{n\times n}/2$

We can see that $\boldsymbol{\theta}^\top(\boldsymbol{I}^\top/2)(\boldsymbol{I}/2)\boldsymbol{\theta} = 1/4\boldsymbol{\theta}^\top\boldsymbol{I}\boldsymbol{I}\boldsymbol{\theta} = 1/4\sum\theta_i^2$

5. (MCQ; 4 marks) Define $\boldsymbol{M}$ such that we obtain the constraint $(\sum\theta_i)^2 < C$.

(a) $[1,1,...,1]^\top{}_{n\times 1}$.
(b) $\boldsymbol{I}_{n\times n}$.
(c) $[1,1,...,1]_{1\times n}$.
(d) $\boldsymbol{I}_{n\times n}/4$.
(e) None of the above.

Correct answers: (c)

Explanation: $M=[1,1,...,1]_{1\times n}$

We can see that $\boldsymbol{\theta}^\top\boldsymbol{M}^\top\boldsymbol{M}\boldsymbol{\theta} = [(\sum\theta_i)]\boldsymbol{M}\boldsymbol{\theta} = [(\sum\theta_i),...,(\sum\theta_i)]\boldsymbol{\theta} = (\sum\theta_i)^2$s