UVA CS 4774: Machine Learning

Part 0: Quiz Reviews

Related:

- QA Part 1 [HERE]
- QA Part 2 [HERE]
- Course Announcement Page [HERE]

Dr. Yanjun Qi

University of Virginia

Department Of Computer Science

Q1

Q1: Given the definitions of A and B below, compute AB.

Multiple choice

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ -1 & 0 \\ 5 & 2 \end{bmatrix}$$

Option 1

$$\begin{bmatrix} -5 & -1 \\ 17 & 8 \end{bmatrix}$$

Option 2

$$\begin{bmatrix} -5 & 1 \\ -17 & 8 \end{bmatrix}$$

Option 3

Q2: For conformable matrices A and B, which of the following always holds?

Multiple choice

- (AB)^T=(B^T)(A^T) ✓
- (AB)^T=AB
- (AB)^T=AB^T
- Add option or add "Other"

Q3:

Multiple choice

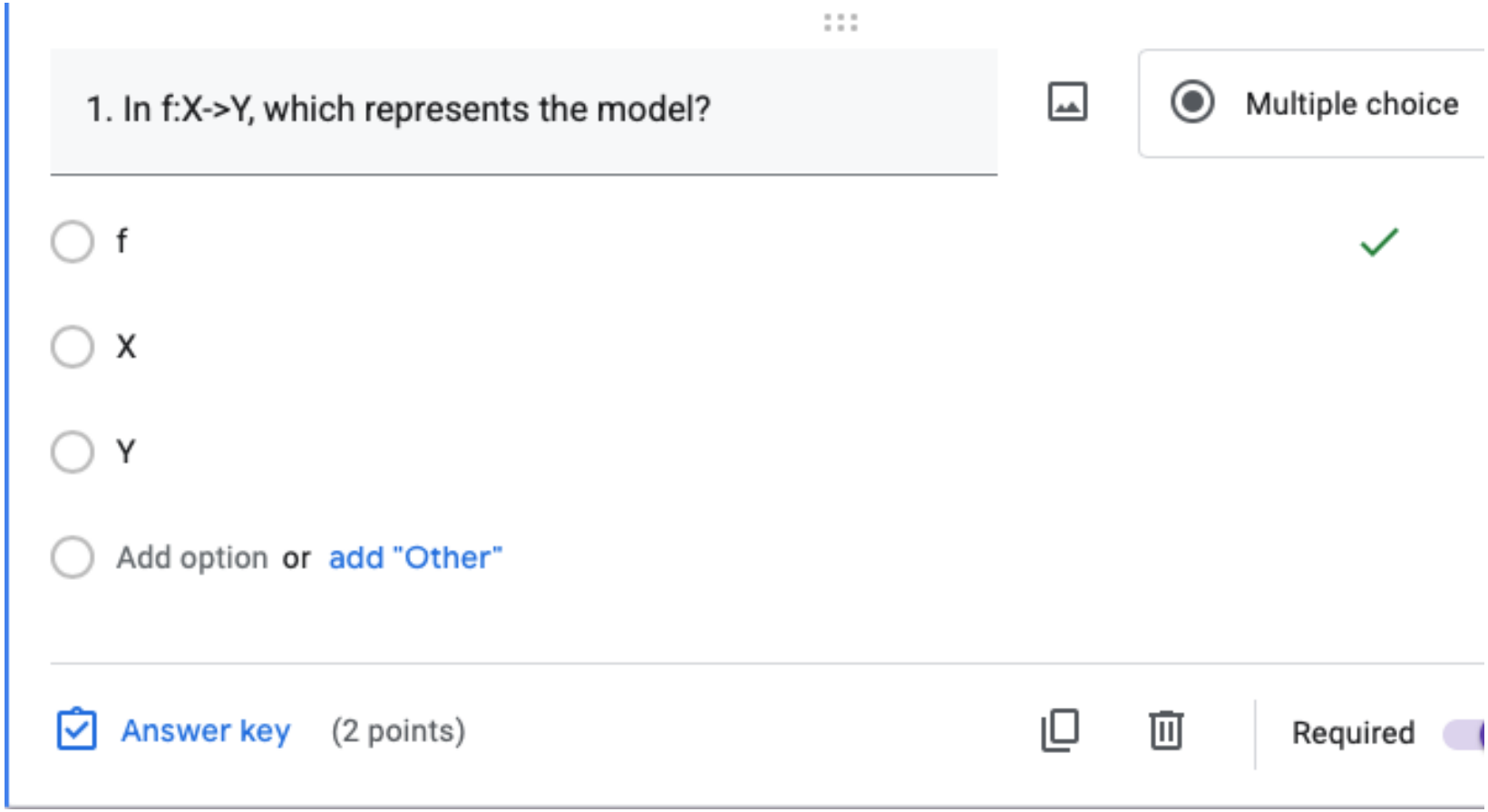
a matrix $D \in R^{5\times 7}$, which of the following must always be true

- Rank(D) = 5
- Rank(D) = 7
- Rank(D) <= 5 ✓
- Rank(D) >= 5
- Rank(D) >= 7
- Add option or add "Other"

Answer key (2 points)

Required

Q2


1. In f:X->Y, which represents the model?
Multiple choice
f
X
Y
Add option or add "Other"
Answer key (2 points)
Required

2. How does unsupervised learning differ from supervised learning?

Multiple choice

- ○ No input X is provided
- ○ No label Y is provided ✓
- ○ Label y is continuous
- ○ Label y is discrete
- ○ Add option or add "Other"

Answer key (2 points)

Required

3. Generalization refers to how well your model performs on

Multiple choice

- [ ] The testing set ✓
- [ ] The training set
- [ ] Both the training and testing set
- [ ] Add option or add "Other"

4. The difference between classification and regression is?

Multiple choice

- Different types of input
- Different types of output ✓
- Different types of model
- Different types of programs
- Add option or add "Other"

Answer key (2 points)

Required

Q3

1. how can we find the best θ that minimizes loss function J(θ) in linear regression? (check all that apply)

Checkboxes

- [ ] Take derivative of J(θ) and set it to 0, solving for θ ✓
- [ ] Calculate the gradient of J(θ) and use gradient descent iteratively ✓
- [ ] Use binary search to find best θ
- [ ] Add option or add "Other"

Answer key (2 points)

Required

2. Suppose loss function J(x) = x^2, the initial value x0=1, learning rate is 0.1, what will be x1 if we apply gradient descent?

Multiple choice

- x1 = -1
- x1 = 0
- x1 = 0.8 ✓
- Add option or add "Other"

Answer key (2 points)

Required

3. Suppose we apply gradient descent with a learning rate that is too large. What is the most likely outcome?

Multiple choice

- Convergence will be faster and guaranteed
- The algorithm may overshoot and fail to converge ✓
- The algorithm will converge to a local minimum instead of a global minimum
- The final solution will always be the closed-form solution
- Add option or add "Other"

Answer key (2 points)

Required

4. Comparing SGD, GD, minibatch-GD, what is the most important difference?

Multiple choice

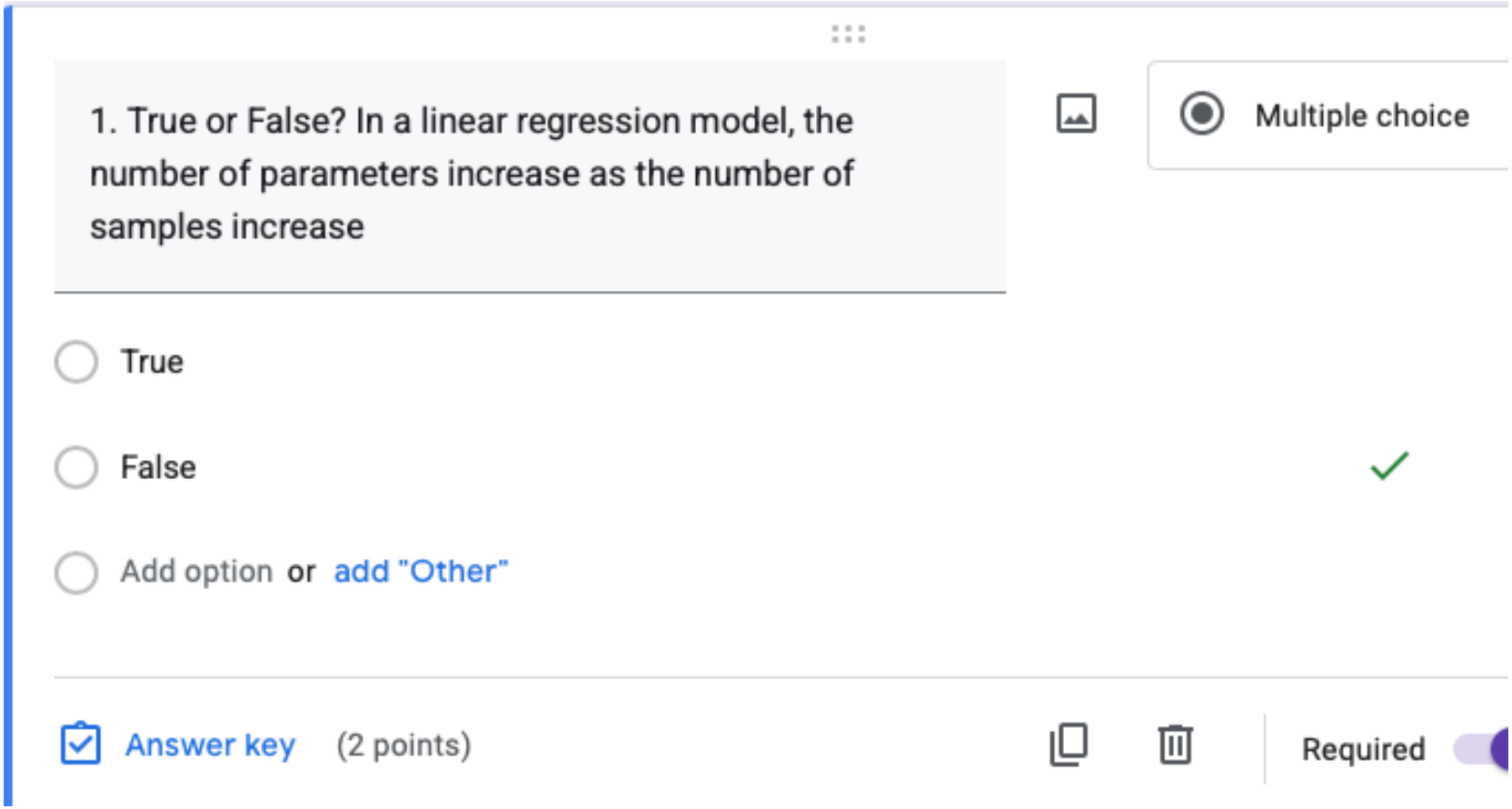
- Error metric/ Loss function
- The number of data used for each update ✓
- Total number of epochs
- Add option or add "Other"

Answer key (2 points)

Required

Q4

Q4 review


1. True or False? In a linear regression model, the number of parameters increase as the number of samples increase
Multiple choice
True
False
Add option or add "Other"
Answer key (2 points)
Required

Q4 review

2. Which data split should the model never be trained on? ()

Multiple choice

- Training Set
- K fold Split
- Test Set
- Add option or add "Other"

Answer key (2 points)

Required

Q4 review

3. Which of the following is an advantage of k-fold cross validation over the simple validation set method for model selection?

Multiple choice

- Requires fewer computations
- Provides a lower-variance estimate of model performance ✓
- Uses less training data overall
- Always achieves lower test error
- Add option or add "Other"

Answer key (2 points)

Required

Q4 review

4. Suppose a model has low training error but very high test error. Which of the following strategies could help? (check all that apply)

Checkboxes

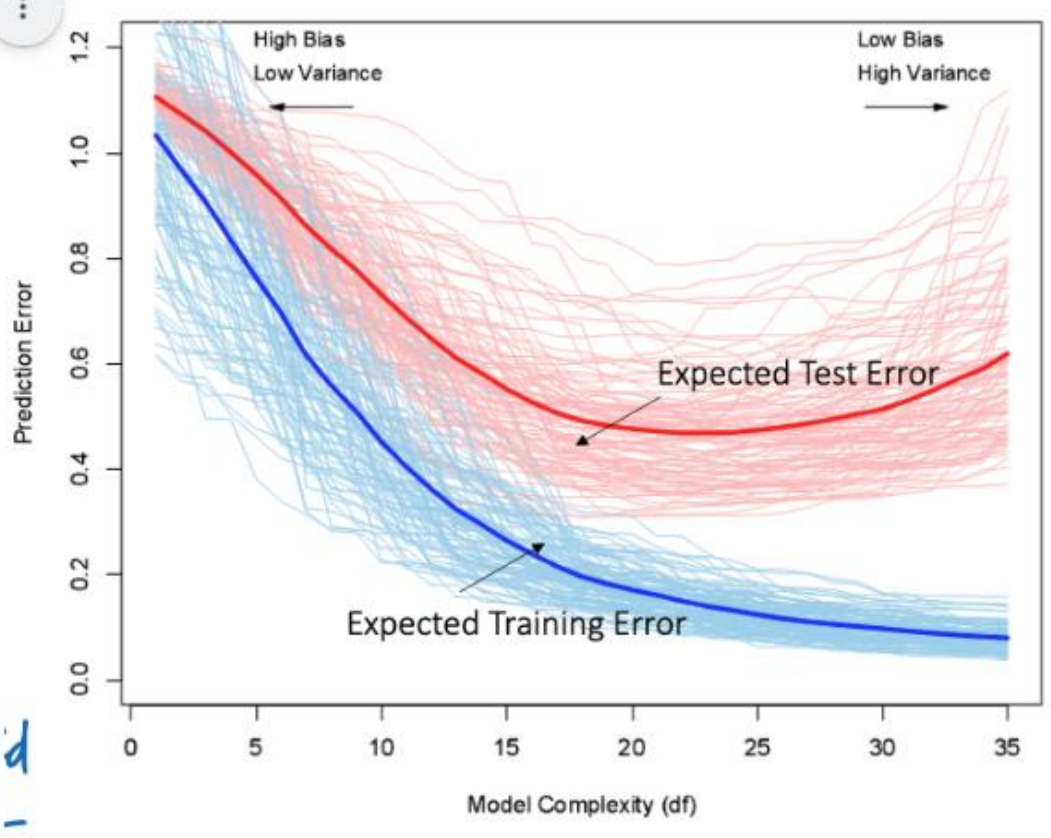
- [ ] Use a more complex learner
- [ ] Add more training data ✓
- [ ] Reduce the number of features ✓
- [ ] Add option or add "Other"

Answer key (2 points)

Required

5. Consider the below diagram. If you have a model with complexity of 5 degrees of freedom (d=5), it will face:

Multiple choice



- Underfitting ✓
- Overfitting
- Neither
- Add option or add "Other"

Q4

Q5 Review

1. Which of the following best explains why Lasso Regression can perform feature selection?

Multiple choice

- It penalizes squared coefficients
- It can set some coefficients exactly to zero ✓
- It increases the training error
- It reduces both training and test errors simultaneously
- Add option or add "Other"

Q5

2. Suppose we scale the weight feature from kilograms to grams in a KNN classifier. What effect does this have if we don't normalize the data?

Multiple choice

- ○ No effect, because KNN is scale-invariant
- ○ Distance measure may get dominated by the weight feature ✓
- ○ The algorithm will automatically adjust feature ranges
- ○ Model complexity will increase
- ○ Add option or add "Other"

Q5

3. Which of the following is a disadvantage of increasing k in KNN when used for classification?

Multiple choice

- Model becomes more sensitive to noise
- Model predictions become less stable
- Model introduces more smooth decision boundary ✓
- The decision boundary becomes more irregular
- Add option or add "Other"

Q5

4. In a vanilla (i.e. regular) k-nearest neighbor classifiers, as k increases, the model

Multiple choice

- becomes more complex and may overfit
- becomes more simple and may underfit ✓
- Add option or add "Other"

Q5

5. Select method that prevent overfitting

Checkboxes

- [ ] Keep increasing model complexity
- [ ] Keep adding training epochs
- [ ] Use proper regularizer ✓
- [ ] Add option or add "Other"

Q5

6. (Extra Credit) In k-nearest neighbor classifiers, suppose k is fixed, as the number of training samples (N) increase, the time it takes to test a new sample:

Multiple choice

- Increases ✓
- Decreases
- Stays the same
- Add option or add "Other"

7. (Extra Credit)

Q5

Checkboxes

Assume we have n training (x, y) pairs, and each x has 5 features. After training with ridge regression loss shown as follows, we get w1 = 0.8, w2 = -1, w3=0.01, w4=0.001, w5=0.1, choose two most important features.

$$L(W) = \frac{1}{n}\sum_{i=1}^{n}[y_i - (w_1x_{i,1} + w_2x_{i,2} + w_3x_{i,3} + w_4x_{i,4} + w_5x_{i,5} + b)]^2 + \beta||W||_2^2$$

- [ ] x1 ✓
- [ ] x2 ✓
- [ ] x3
- [ ] x4
- [ ] x5
- [ ] Add option or add "Other"

Q6

1. In L2-regularized linear regression, increasing the regularization strength λ generally...

Multiple choice

- decreases bias and increases variance
- increases bias and decreases variance ✓
- decreases both bias and variance
- increases both bias and variance
- Add option or add "Other"

Answer key (2 points)

Required

2. For k-NN on a fixed dataset, which k is most vulnerable to overfitting and what happens to the training error?

Multiple choice

- k=1; training error ≈ 0 ✓
- k=15; training error ≈ 0
- k=15; training error very high
- k=1; training error very high
- Add option or add "Other"

Answer key (2 points)

Required

3. Which actions are appropriate when a model shows high variance? (select all that apply)

Checkboxes

- [x] Use a simpler model
- [x] Add regularization
- [x] Get more training data
- [x] Reduce feature set
- [x] Build an ensemble (bagging)
- [ ] Add option or add "Other"

Answer key (2 points)

Required

4. In the diagram below, the red center shows the true values, and the blue points show the predictions. The model has:

Multiple choice

- [x] High bias, low varience
- [ ] High bias, high varience
- [ ] Low bias, low varience
- [ ] Low bias, high varience
- [ ] Add option or add "Other"

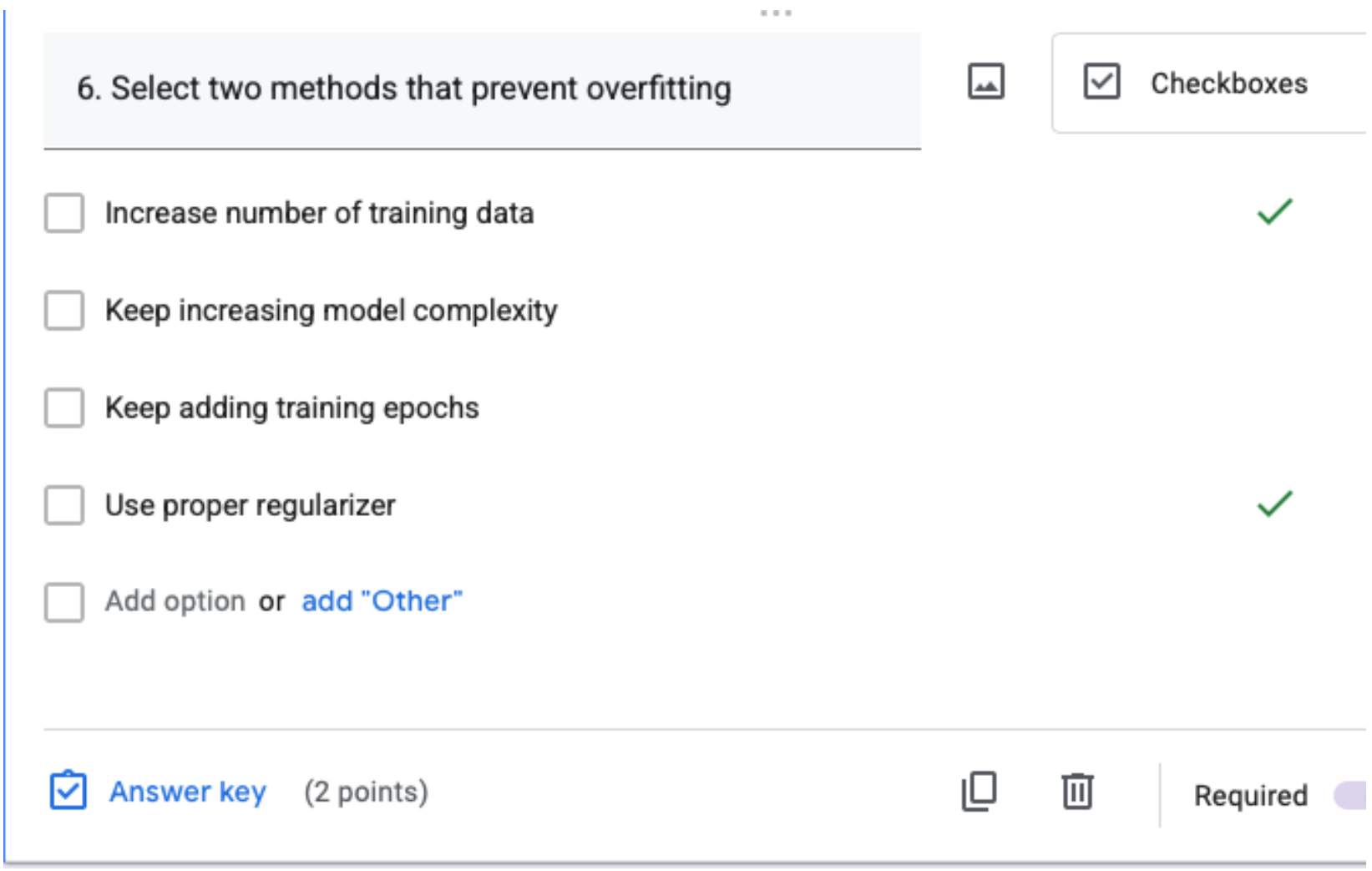
5. In a vanilla (i.e. regular) k-nearest neighbor classifiers, as k increases, the model

Multiple choice

- becomes more complex and may overfit
- becomes more simple and may underfit ✓
- Add option or add "Other"

Answer key (2 points)

Required


6. Select two methods that prevent overfitting
Checkboxes
Increase number of training data
Keep increasing model complexity
Keep adding training epochs
Use proper regularizer
Add option or add "Other"
Answer key (2 points)
Required

7. If A and B are independent random variables, then P(A and B) =

Multiple choice

- P(A)*P(B) ✓
- P(A) + P(B)
- Add option or add "Other"

Answer key (2 points)

Required

Q7

1. Logistic regression is considered a discriminative model because it:

Multiple choice

- Models the joint probability P(X,Y) directly.
- Models the conditional probability P(Y|X) directly. ✓
- Assumes features are independent given the class label.
- Uses Bayes' theorem to estimate priors and likelihoods.
- Add option or add "Other"

Answer key (2 points)

Required

2. At the decision boundary of logistic regression, P(y = 1 | x) = P(y = 0 | x) = ?

Multiple choice

- 1
- -1
- 0.5 ✓
- 0

Add option or add "Other"

Answer key (2 points)

Required

3. True or False? Logistic regression models have a linear decision boundary.

Multiple choice

- True ✓
- False
- Add option or add "Other"

Answer key (2 points)

Required

4: Logistic regression is:

Multiple choice

- An unsupervised algorithm
- A supervised algorithm ✓
- Add option or add "Other"

Answer key (2 points)

Required

5. The decision boundary of logistic regression is:

Multiple choice

- Always linear in feature space ✓
- Always nonlinear in feature space
- Quadratic in feature space
- Randomly determined by initialization
- Add option or add "Other"

Answer key (2 points)

Required

6. In a neural network, the activation function introduces:

Multiple choice

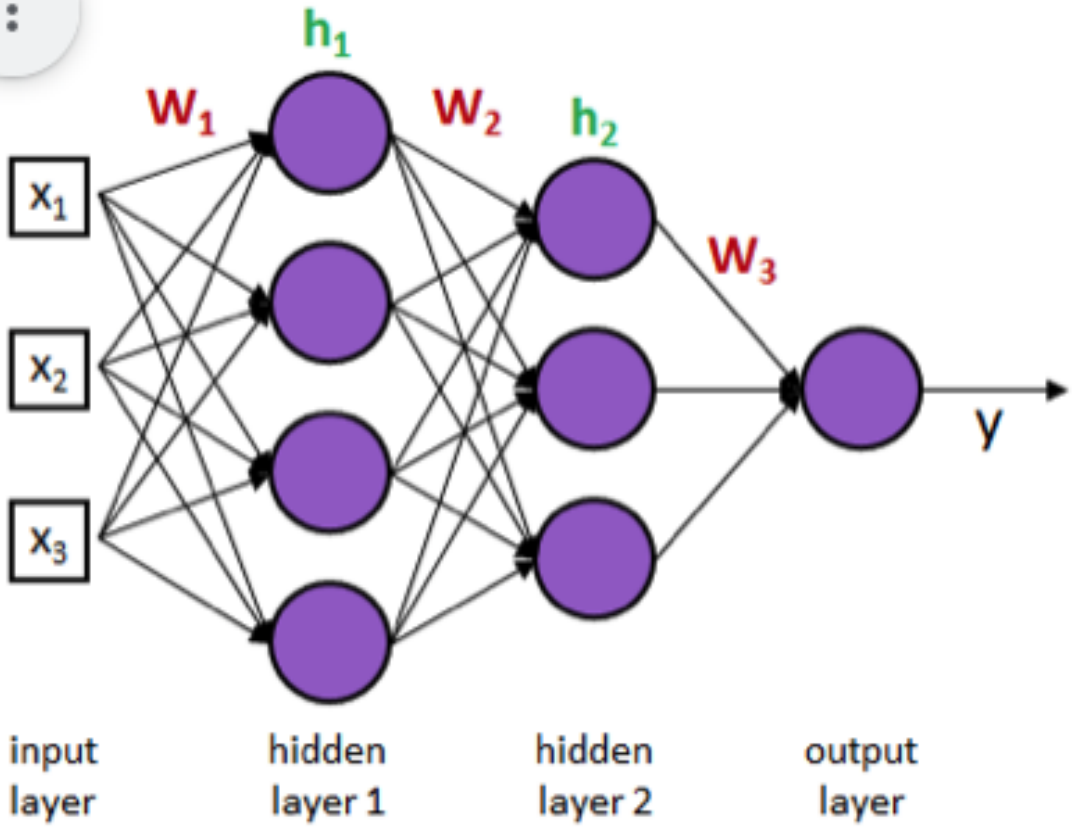
- Linearity
- Non-linearity ✓
- Bias
- Regularization
- Add option or add "Other"

Answer key (2 points)

Required

Multiple choice

7. Ignoring bias terms, what is the dimension of the weight matrix W2 in the following neural network?



3 x 3

8. True or False? Summation of the outputs of a softmax function is 1.

Multiple choice

- True ✓
- False
- Add option or add "Other"

Answer key (2 points)

Required

Q8