GO CLASSES

More Questions on CV

Question: 22

v. (1 point) Suppose we use leave-one-out cross validation meaning we use 7-fold cross validation with a split of 6 to 1 between training set and validating set. Compute the average classification error over the 7-folds. (Hard Margin SVM)

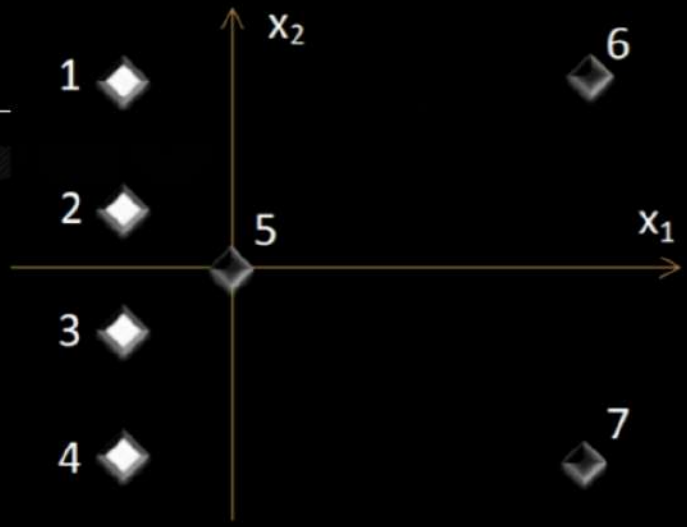


Figure 5: Training Data

https://nyu-ds1003.github.io/mlcourse/2021/exams/sp20/midterm_solutions.pdf

Figure 5: Training Data

$\frac{1}{7}$

7-fold cross validation

misclassified

1 (val)	2 3 4 5 6 7 (train)	0
2 (val)	1 3 4 5 6 7 (train)	0
5 (val)	1 2 3 4 6 7 (train)	1

v. (1 point) Suppose we use leave-one-out cross validation meaning we use 7-fold cross validation with a split of 6 to 1 between training set and validating set. Compute the average classification error over the 7-folds.

Solution: 1/7

https://nyu-ds1003.github.io/mlcourse/2021/exams/sp20/midterm_solutions.pdf

GO CLASSES

leave one out cross validation

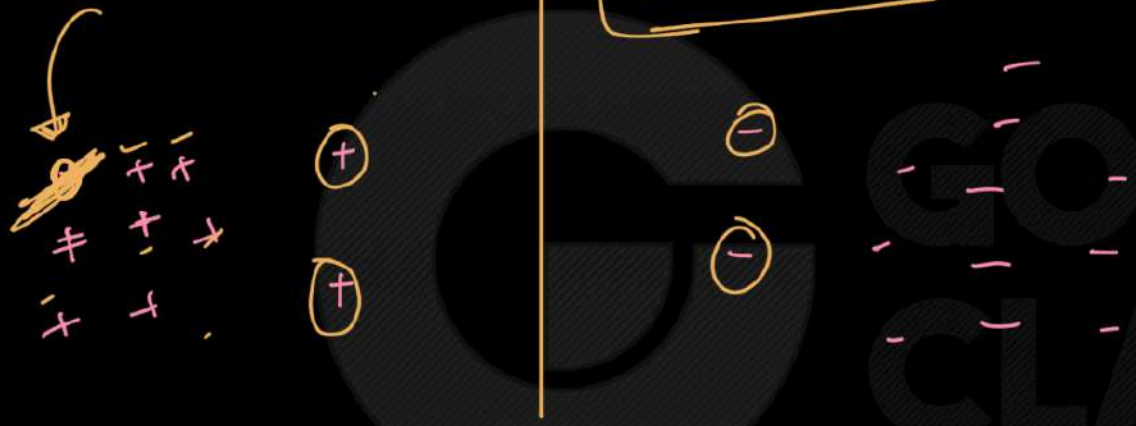
$$w = \sum \alpha_i x_i y_i$$

if you remove the "non-support" vector and train then put it again (as validation) then it will always be $\Rightarrow$ correctly classified.

GO CLASSES

leave one out cross validation

$$w = \sum \alpha_i x_i y_i$$



if you remove the "non-support" vector and train then put it again (as validation) then it will always be $\Rightarrow$ correctly classified.

No. of support vectors $\exists$ 10

Suppose there are 10^5 points in training and 10 are Support vectors

Now suppose we use LOOCV then what is maximum number of misclassification possible?

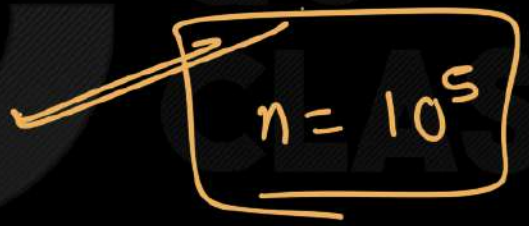
$\frac{10}{10^5}$

Question: 23 True/False

(19) True/False: k-fold cross-validation with $k = 100$ is computationally more expensive (slower) than "leave-one-out" cross validation. (Assume that there are enough data points to divide the dataset evenly by k.)

(A) True

(B) False



https://courses.cs.washington.edu/courses/cse446/23au/exams/pastexams/22au-midterm-solutions.pdf

(19) True/False: k-fold cross-validation with $k = 100$ is computationally more expensive (slower) than "leave-one-out" cross validation. (Assume that there are enough data points to divide the dataset evenly by k.)

(A) True

(B) False

Solution:

The solution is (B)

Question: 24

(20) Assume we have a data matrix X. Which of the following is a true statement when comparing leave-one-out cross validation (LOOCV) error with the true error?

(A) LOOCV error is typically a slight underestimation of the true error of a model trained on X.

(B) LOOCV error is typically a slight overestimation of the true error of a model trained on X.

(C) LOOCV error is an unbiased estimator of the true error of a model trained on X.

https://courses.cs.washington.edu/courses/cse446/23au/exams/pastexams/22au-midterm-solutions.pdf

(20) Assume we have a data matrix X. Which of the following is a true statement when comparing leave-one-out cross validation (LOOCV) error with the true error?

(A) LOOCV error is typically a slight underestimation of the true error of a model trained on X.

(B) LOOCV error is typically a slight overestimation of the true error of a model trained on X.

(C) LOOCV error is an unbiased estimator of the true error of a model trained on X.

Solution:

The solution is (B)

Question: 25

9. Leave-one-out cross-validation (LOOCV) is a special case of k-fold cross-validation where:

(a) The training set contains all but one sample, and the remaining sample is used for testing.

(b) The training set contains only one sample, and the remaining sample is used for testing.

(c) The training set contains exactly one sample from each class, and the remaining samples are used for testing.

(d) The training set contains one sample from each fold, and the remaining folds are used for testing.

9. Leave-one-out cross-validation (LOOCV) is a special case of k-fold cross-validation where:

(a) The training set contains all but one sample, and the remaining sample is used for testing.

(b) The training set contains only one sample, and the remaining sample is used for testing.

(c) The training set contains exactly one sample from each class, and the remaining samples are used for testing.

(d) The training set contains one sample from each fold, and the remaining folds are used for testing.

Correct answers: (a)

Question: 26

(a) (4.0 points)

Recall that a random forest is created from a number of decision trees, with each decision tree created from a bootstrapped version of the original training set. One hyperparameter of a random forest is the number of decision trees we train to create the random forest.

Define T to be the number of decision trees used to create the random forest. Let's say we have two candidate values for T: $var1$ and $var2$. We want to perform $var3$ - fold cross-validation to determine the optimal value of T. Assume $var1$, $var2$, and $var3$ are integers.

Some hyper param.

models

T – 2 ← hyper param.

var1, var2

var3 - fold cross val.

i. (2.0 pt) In this cross-validation process, how many **random forests** will we train? Your answer should be in terms of $var1$, $var2$, and/or $var3$ and should be an integer.

ii. (2.0 pt) In this cross-validation process, how many **decision trees** will we train? Your answer should be in terms of $var1$, $var2$, and/or $var3$ and should be an integer.

Question: 26

(a) (4.0 points)

Recall that a random forest is created from a number of decision trees, with each decision tree created from a bootstrapped version of the original training set. One hyperparameter of a random forest is the number of decision trees we train to create the random forest.

Define T to be the number of decision trees used to create the random forest. Let's say we have two candidate values for T: $var1$ and $var2$. We want to perform $var3$ - fold cross-validation to determine the optimal value of T. Assume $var1$, $var2$, and $var3$ are integers.

Some hyper param.

models

T – ? ← hyper param.

var1, var2

var3 - fold cross val.

i. (2.0 pt) In this cross-validation process, how many ~~**random forests**~~ will we train? Your answer should be in terms of $var1$, $var2$, and/or $var3$ and should be an integer.

2 * var3

ii. (2.0 pt) In this cross-validation process, how many **decision trees** will we train? Your answer should be in terms of $var1$, $var2$, and/or $var3$ and should be an integer.

this is related to Random forest so skipping this ques. as Random forest is NOT in syllabus.

(a) (4.0 points)

Recall that a random forest is created from a number of decision trees, with each decision tree created from a bootstrapped version of the original training set. One hyperparameter of a random forest is the number of decision trees we train to create the random forest.

Define T to be the number of decision trees used to create the random forest. Let's say we have two candidate values for T: $var1$ and $var2$. We want to perform $var3$ - fold cross-validation to determine the optimal value of T. Assume $var1$, $var2$, and $var3$ are integers.

i. **(2.0 pt)** In this cross-validation process, how many **random forests** will we train? Your answer should be in terms of $var1$, $var2$, and/or $var3$ and should be an integer.

2 * var3

ii. **(2.0 pt)** In this cross-validation process, how many **decision trees** will we train? Your answer should be in terms of $var1$, $var2$, and/or $var3$ and should be an integer.

(var1 + var2) * var3

Question: 27

5. Suppose we have a design matrix $\mathbb{X}$ comprising of n observations, d features, and an additional intercept term. We decide to use $\mathbb{X}$ to create, tune, and evaluate a regularized linear regression model with two regularization hyperparameters, λ_1 and λ_2. We have i different choices for λ_1 and j different choices for λ_2, so there are $i \cdot j$ possible combinations of λ_1 and λ_2.

   We set aside 20% of our data to use as a test set and perform k-fold cross-validation to tune our hyperparameters. We compute the cross-validation error of each of the $i \cdot j$ possible combinations of λ_1 and λ_2 values, and our goal is to find the combination of values with the lowest cross-validation error. All following answers should be expressed in terms of i, j, n, d, k, and/or constants only (except for part c).

   (a) [2 Pts] For a single combination of hyperparameter values, how many model parameters do we fit?

   (b) [2 Pts] How many observations are in the validation set of each fold?

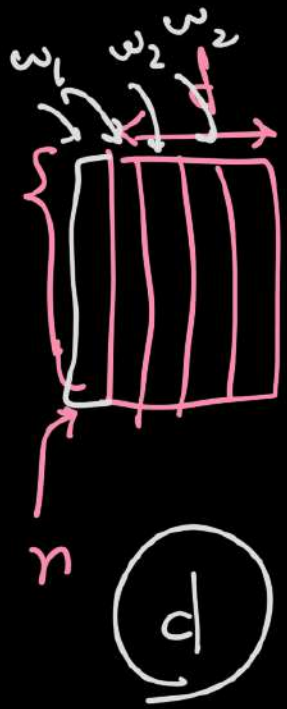
   (c) [2 Pts] How many model parameters do we calculate in total? Your answer can be in terms of A, your answer to part a.

for λ_1

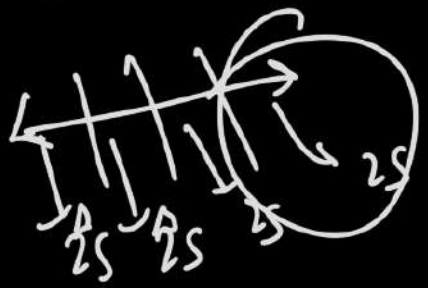
for λ_2

train model

d+1 parameters



Question: 27



for λ_1

for λ_2

5. Suppose we have a design matrix $\mathbb{X}$ comprising of n observations, d features, and an additional intercept term. We decide to use $\mathbb{X}$ to create, tune, and evaluate a regularized linear regression model with two regularization hyperparameters, λ_1 and λ_2. We have i different choices for λ_1 and j different choices for λ_2, so there are $i \cdot j$ possible combinations of λ_1 and λ_2.

   We set aside 20% of our data to use as a test set and perform k-fold cross-validation to tune our hyperparameters. We compute the cross-validation error of each of the $i \cdot j$ possible combinations of λ_1 and λ_2 values, and our goal is to find the combination of values with the lowest cross-validation error. All following answers should be expressed in terms of i, j, n, d, k, and/or constants only (except for part c).

   (a) [2 Pts] For a single combination of hyperparameter values, how many model parameters do we fit?

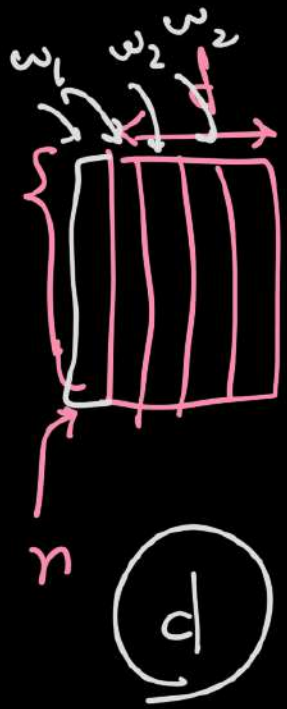
   $K(d+1)$

   (b) [2 Pts] How many observations are in the validation set of each fold?

   $\frac{0.8n}{K}$

   (c) [2 Pts] How many model parameters do we calculate in total? Your answer can be in terms of A, your answer to part a.

   $i \cdot j \cdot K(d+1)$



5. Suppose we have a design matrix $\mathbb{X}$ comprising of n observations, d features, and an additional intercept term. We decide to use $\mathbb{X}$ to create, tune, and evaluate a regularized linear regression model with two regularization hyperparameters, λ_1 and λ_2. We have i different choices for λ_1 and j different choices for λ_2, so there are $i \cdot j$ possible combinations of λ_1 and λ_2.

   We set aside 20% of our data to use as a test set and perform k-fold cross-validation to tune our hyperparameters. We compute the cross-validation error of each of the $i \cdot j$ possible combinations of λ_1 and λ_2 values, and our goal is to find the combination of values with the lowest cross-validation error. All following answers should be expressed in terms of i, j, n, d, k, and/or constants only (except for part c).

   (a) [2 Pts] For a single combination of hyperparameter values, how many model parameters do we fit?

   Solution: $k \cdot (d+1)$

   (b) [2 Pts] How many observations are in the validation set of each fold?

   Solution: $\frac{0.8n}{k}$

   (c) [2 Pts] How many model parameters do we calculate in total? Your answer can be in terms of A, your answer to part a.

   Solution: $i \cdot j \cdot k \cdot (d+1)$

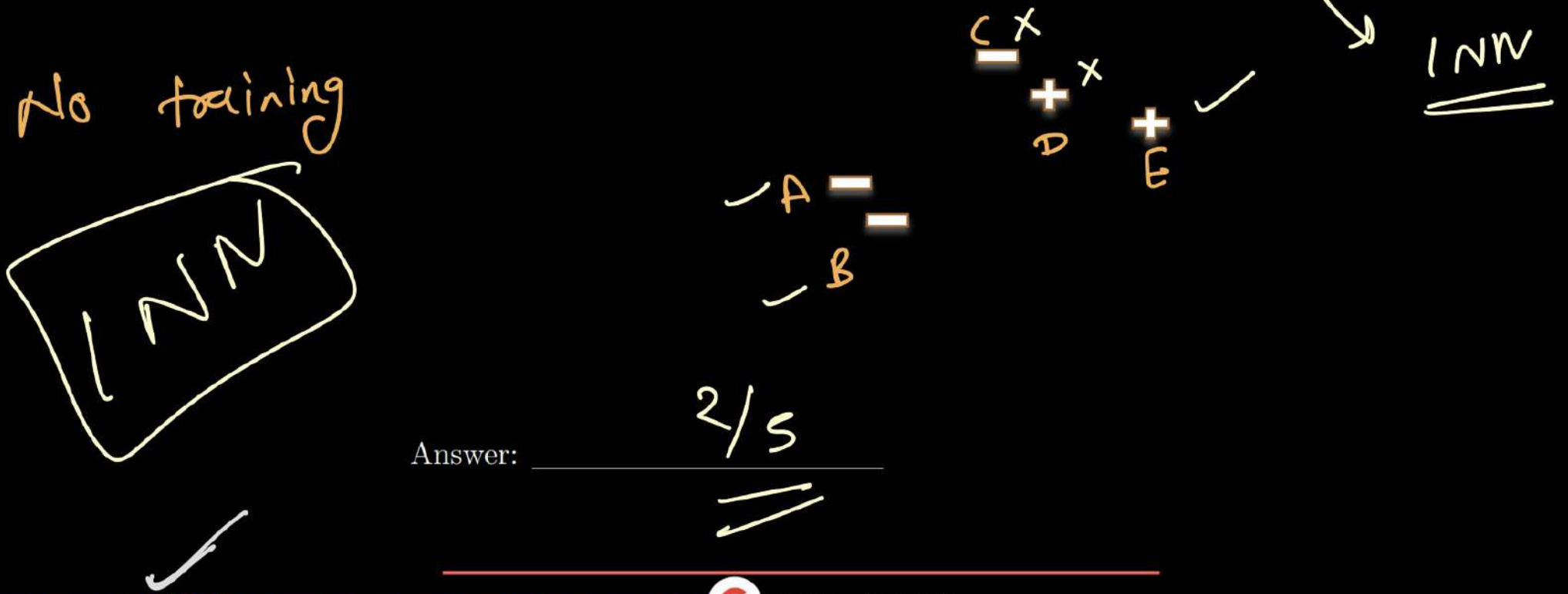
Question:

35. **Extra credit:** Consider a nearest neighbor classifier that chooses the label for a test point to be the label of its nearest neighboring training example. What is its leave-one-out cross-validated error for the data in the following figure? ("+" and "−" indicate labels of the points).

Answer: _____

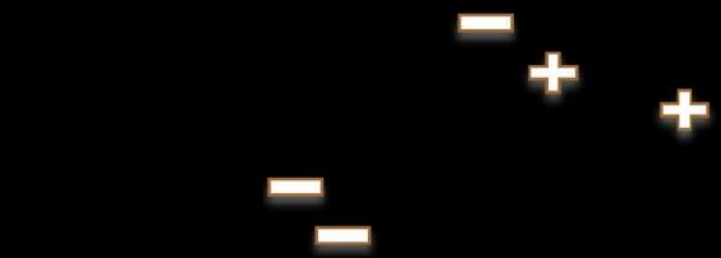
Question:

35. **Extra credit:** Consider a nearest neighbor classifier that chooses the label for a test point to be the label of its nearest neighboring training example. What is its leave-one-out cross-validated error for the data in the following figure? ("+" and "−" indicate labels of the points).



Answer: 2/5

35. **Extra credit:** Consider a nearest neighbor classifier that chooses the label for a test point to be the label of its nearest neighboring training example. What is its leave-one-out cross-validated error for the data in the following figure? ("+" and "−" indicate labels of the points).



5 mins

2-3mins

Answer: _____

Explanation: The solution is 2/5

Question:

4×3

15. [2.5 Pts] Aman and Ed built a model on their data with two regularization hyperparameters λ and γ. They have 4 good candidate values for λ and 3 possible values for γ, and they are wondering which λ, γ pair will be the best choice. If they were to perform five-fold cross-validation, how many validation errors would they need to calculate?

https://ds100.org/fa20/resources/assets/exams/fa18/fa18finalsol.pdf

15. [2.5 Pts] Aman and Ed built a model on their data with two regularization hyperparameters λ and γ. They have 4 good candidate values for λ and 3 possible values for γ, and they are wondering which λ, γ pair will be the best choice. If they were to perform five-fold cross-validation, how many validation errors would they need to calculate?

easy question

Solution: 60 $4 \times 3 \times 5 = 60$

for each possible value of $\lambda, \gamma \Rightarrow$ we need to train 5 times each

Question:

13. Why is it important to use a different test set to evaluate the final performance of the model, rather than the validation set used during model selection?

(a) The model may have overfit to the validation set

(b) The test set is a better representation of new, unseen data

(c) Both a and b

(d) None of the above

https://courses.cs.washington.edu/courses/cse446/23au/exams/pastexams/23sp-midterm-solutions.pdf

13. Why is it important to use a different test set to evaluate the final performance of the model, rather than the validation set used during model selection?

(a) The model may have overfit to the validation set

(b) The test set is a better representation of new, unseen data

(c) Both a and b

(d) None of the above

Correct answers: (c)

MSQ Question:

(g) [4 pts] You are training a soft-margin SVM on a binary classification problem. You find that your model's training accuracy is very high, while your validation accuracy is very low. Which of the following are likely to improve your model's performance on the validation data?

- ○ A: Training your model on more data.
- ○ B: Adding a quadratic feature to each sample point.
- ○ C: Increasing the hyperparameter C.
- ○ D: Decreasing the hyperparameter C.

typical example of overfit

https://people.eecs.berkeley.edu/~jrs/189/exam/mids22.pdf

MSQ Question:

(g) [4 pts] You are training a soft-margin SVM on a binary classification problem. You find that your model's training accuracy is very high, while your validation accuracy is very low. Which of the following are likely to improve your model's performance on the validation data?

- ○ A: Training your model on more data.
- ○ B: Adding a quadratic feature to each sample point.
- ○ C: Increasing the hyperparameter C.
- ○ D: Decreasing the hyperparameter C.

typical example of overfit

Hard marg. ⇒ overfit

Complex

we want to simplify the model & it will not help me

https://people.eecs.berkeley.edu/~jrs/189/exam/mids22.pdf

(g) [4 pts] You are training a soft-margin SVM on a binary classification problem. You find that your model's training accuracy is very high, while your validation accuracy is very low. Which of the following are likely to improve your model's performance on the validation data?

- ● A: Training your model on more data.
- ○ B: Adding a quadratic feature to each sample point.
- ○ C: Increasing the hyperparameter C.
- ● D: Decreasing the hyperparameter C.

A is true as training on more data reduces variance and decreases overfitting in general. B is false since polynomial features increase the variance and the risk of overfitting. C is false because increasing C enforces a harder margin constraint, leading to more overfitting. D is true because decreasing C allows for more slack, decreasing the risk of overfitting.

→ Cross valid

→ precision recall

→ Bias variance trade-off

MSQ Question:

Which of the following statement(s) is(are) true for k-NN?

- ☐ k represents the number of classes.
- ☐ The final outcome of the algorithm may change with the distance measure.
- ☐ CrossValidation can be used to find the optimal k.
- ☐ As k increases, we overfit the data eventually.

https://moodlearchive.epfl.ch/2022-2023/pluginfile.php/3076586/mod_resource/content/1/Mock_Exam_2022_Spring_Solutions.pdf

MSQ Question:

Which of the following statement(s) is(are) true for k-NN?

- ☐ k represents the number of classes.
- ☐ The final outcome of the algorithm may change with the distance measure.
- ☐ CrossValidation can be used to find the optimal k.
- ☐ As k increases, we overfit the data eventually.

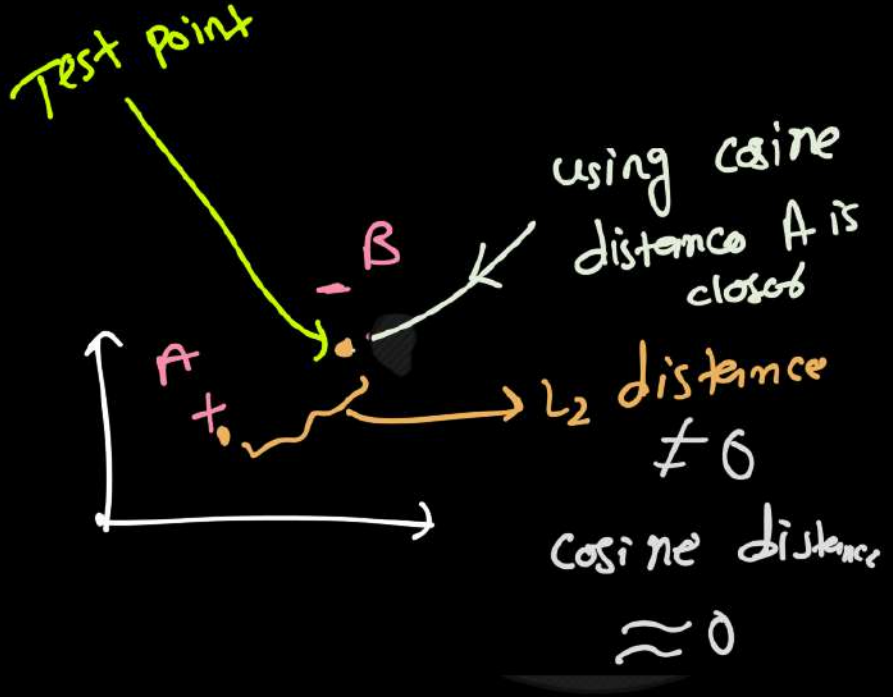
it underfit

30 examples ⇒ +ve
20 ⇒ -ve

50 examples
$k = 50$

using cosine distance A is closer

cosine distance
≈ 0



for the test point

- using cosine distance A is closer so classified as "+"
- using L_2 distance B is closer so classified as "−".

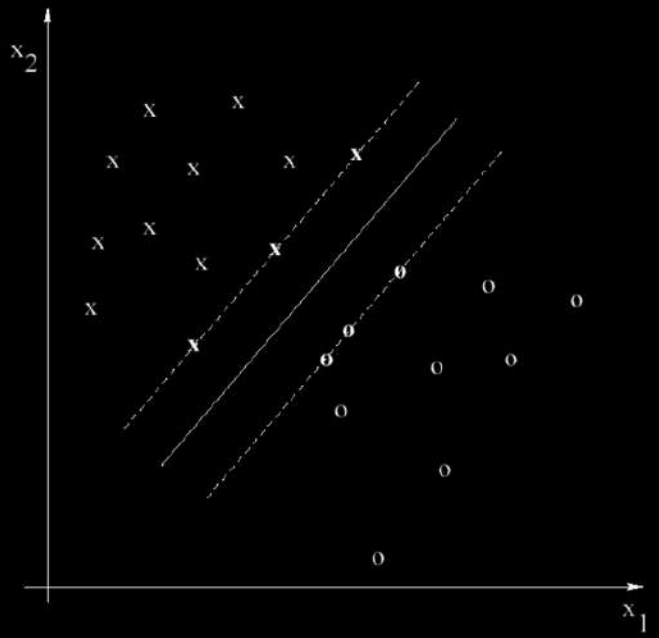
Which of the following statement(s) is(are) true for k-NN?

- ☐ k represents the number of classes.
- ☒ The final outcome of the algorithm may change with the distance measure.
- ☒ CrossValidation can be used to find the optimal k.
- ☐ As k increases, we overfit the data eventually.

Question:

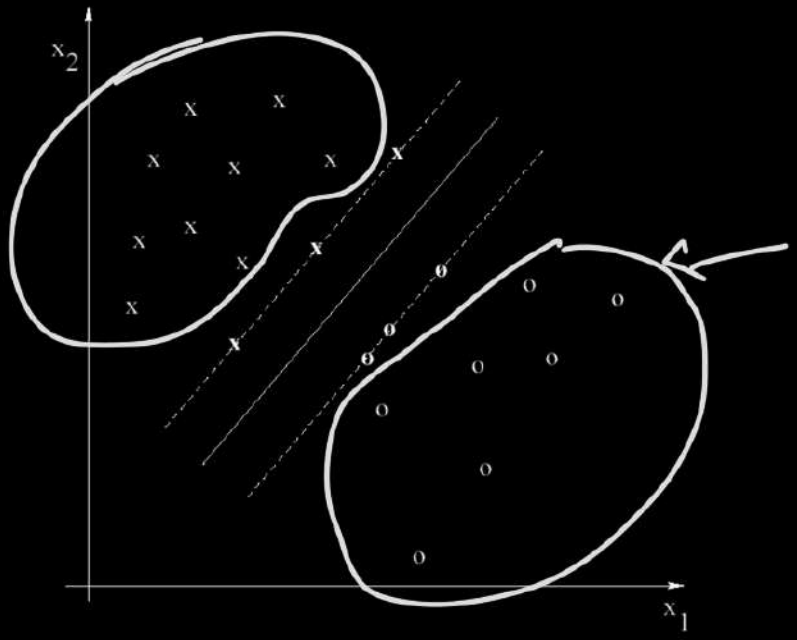
What is the leave-one-out cross-validation error estimate for maximum margin separation in the following figure ? (we are asking for a number)

SVM



http://www.ai.mit.edu/courses/6.867-f03/exams/midterm2001sols.pdf

What is the leave-one-out cross-validation error estimate for maximum margin separation in the following figure ? (we are asking for a number)



we don't need to care about these

Answer: 0

Based on the figure we can see that removing any single point would not chance the resulting maximum margin separator. Since all the points are initially classified correctly, the leave-one-out error is zero.

What do you do after k-fold cross validation.

- Once you have decided which model or set of parameters to use, you then train a new model over the whole data set and use that for prediction.

Question:

[2 points] Suppose you have picked the parameter θ for a model using 10-fold cross validation. The best way to pick a final model to use and estimate its error is to

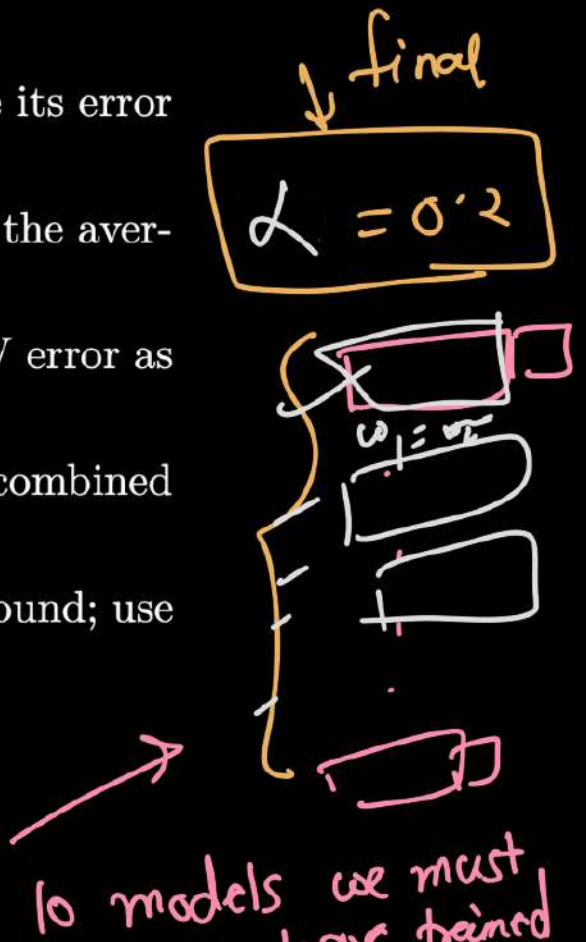
(a) pick any of the 10 models you built for your model; use its error estimate on the held-out data

(b) pick any of the 10 models you built for your model; use the average CV error for the 10 models as its error estimate

(c) average all of the 10 models you got; use the average CV error as its error estimate

(d) average all of the 10 models you got; use the error the combined model gives on the full training set

(e) train a new model on the full data set, using the θ you found; use the average CV error as its error estimate



[2 points] Suppose you have picked the parameter θ for a model using 10-fold cross validation. The best way to pick a final model to use and estimate its error is to

(a) pick any of the 10 models you built for your model; use its error estimate on the held-out data

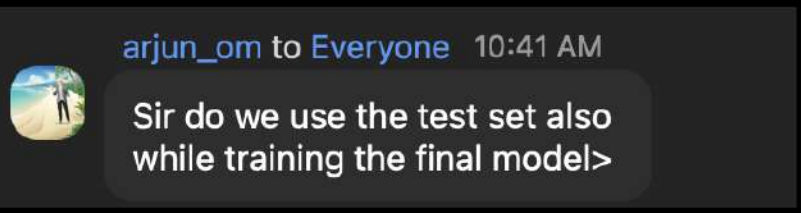
(b) pick any of the 10 models you built for your model; use the average CV error for the 10 models as its error estimate

(c) average all of the 10 models you got; use the average CV error as its error estimate

(d) average all of the 10 models you got; use the error the combined model gives on the full training set

(e) train a new model on the full data set, using the θ you found; use the average CV error as its error estimate

★ **SOLUTION:** E



Question:

[3] Under which of the following conditions is ***k*-fold cross-validation** the *same as* **leave-one-out cross-validation**?

A. The training set and test set have the *same* number of examples
B. The training set and tuning set have the *same* number of examples
C. $k = 1$
D. $k = n$, where n is the total number of examples
E. None of the above

[3] Under which of the following conditions is ***k*-fold cross-validation** the *same as* **leave-one-out cross-validation**?

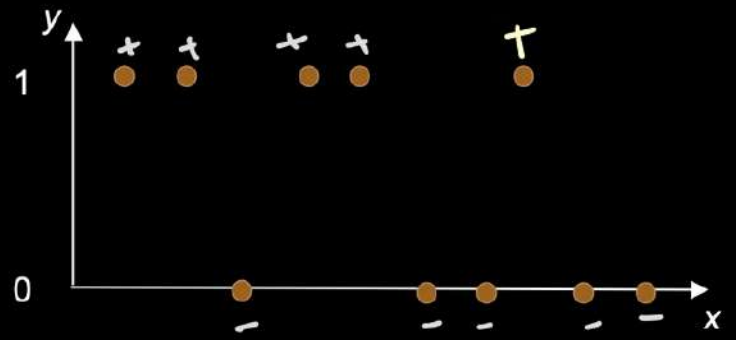
A. The training set and test set have the *same* number of examples
B. The training set and tuning set have the *same* number of examples
C. $k = 1$
D. $k = n$, where n is the total number of examples
E. None of the above

D is Correct

Question:

Suppose you are using a Majority Classifier on the following training set containing 10 examples where each example has one real-valued feature, *x*, and a binary class label, y, with value 0 or 1. Define this Majority Classifier to predict the class label that is in the *majority in the training set*, regardless of the input value. In case of ties, predict class 1.

LOOCV



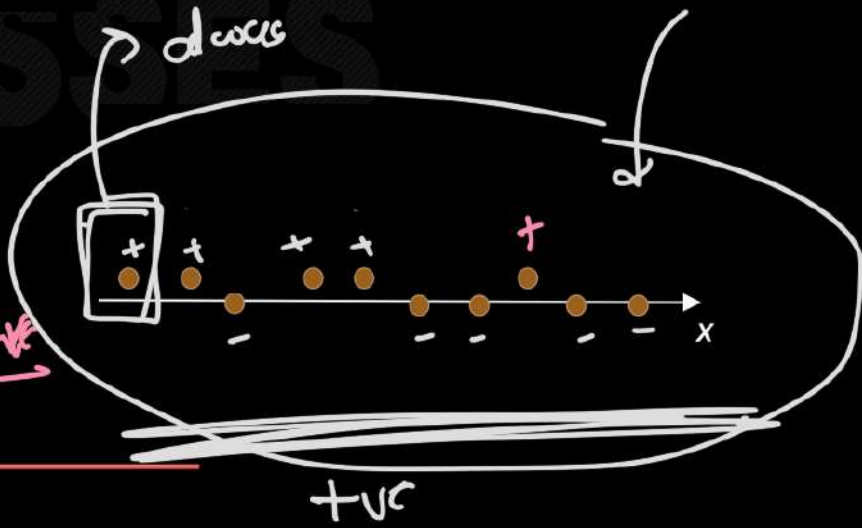
total ⇒ 5 +ve
total ⇒ 5 -ve

(a) [3] What is the *training set accuracy*?

(i) 0%
(ii) 10%
(iii) 50%
(iv) 90%
(v) 100%

$$= \sum (y - w^T x)^2$$

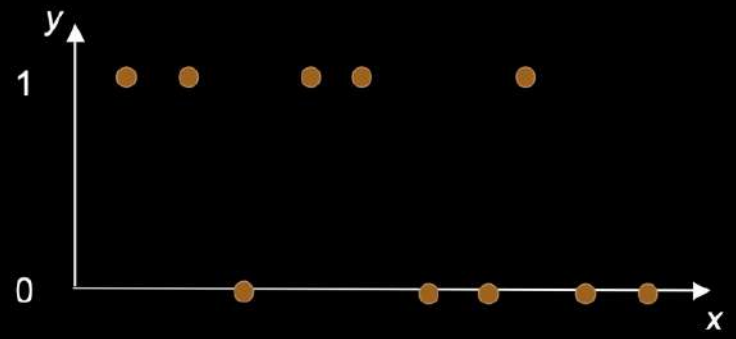
1so-NN



in the prediction, every point is +ve

https://pages.cs.wisc.edu/~dyer/cs540/exams/exam1_f18-sol.pdf

Suppose you are using a Majority Classifier on the following training set containing 10 examples where each example has one real-valued feature, *x*, and a binary class label, y, with value 0 or 1. Define this Majority Classifier to predict the class label that is in the *majority in the training set*, regardless of the input value. In case of ties, predict class 1.



(a) [3] What is the *training set accuracy*?

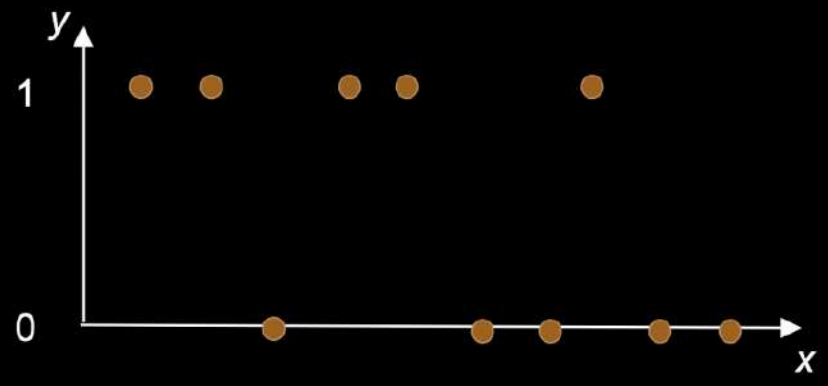
(i) 0%
(ii) 10%
(iii) 50%
(iv) 90%
(v) 100%

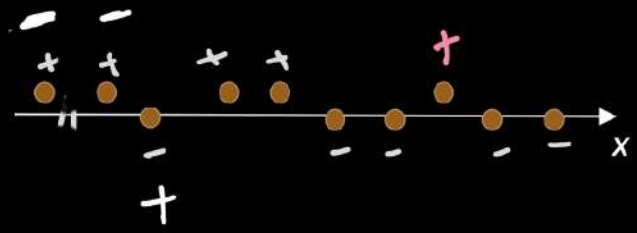
(iii) 5/10 = 50%

Question:

(b) [3] What is the *Leave-1-Out Cross-Validation accuracy*?

(i) 0%
(ii) 10%
(iii) 50%
(iv) 90%
(v) 100%



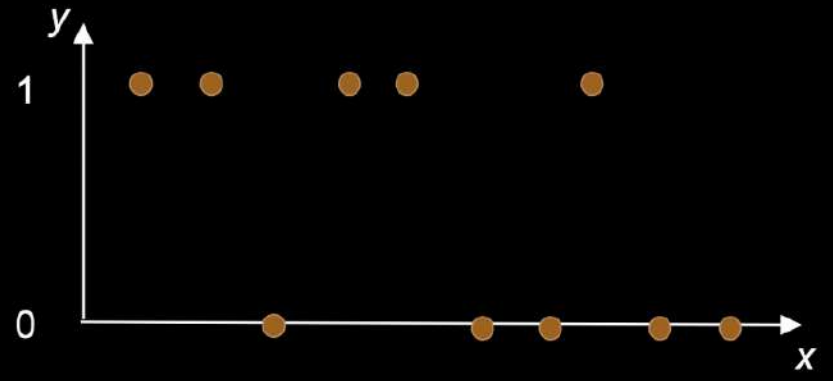


https://pages.cs.wisc.edu/~dyer/cs540/exams/exam1_f18-sol.pdf

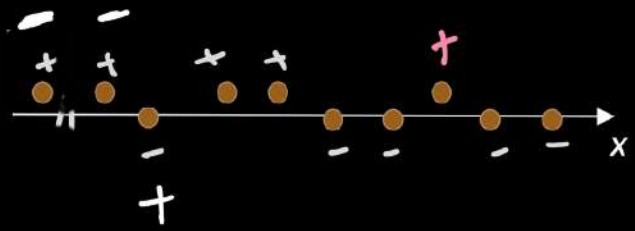
Question:

(b) [3] What is the *Leave-1-Out Cross-Validation accuracy*?

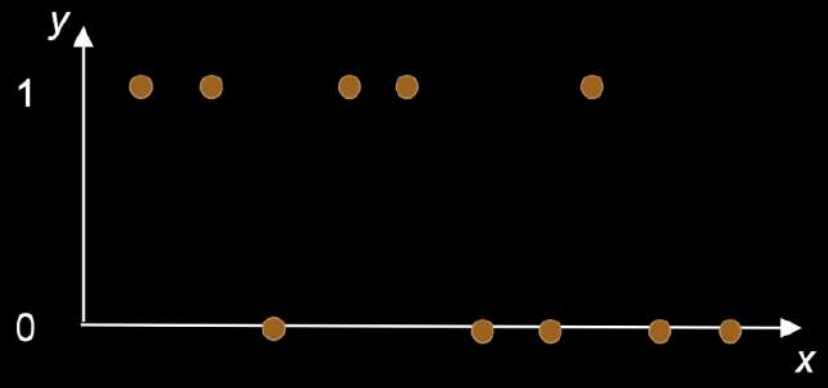
(i) 0%
(ii) 10%
(iii) 50%
(iv) 90%
(v) 100%



here every point will be misclassified



(b) [3] What is the *Leave-1-Out Cross-Validation accuracy*?



(i) 0%
(ii) 10%
(iii) 50%
(iv) 90%
(v) 100%

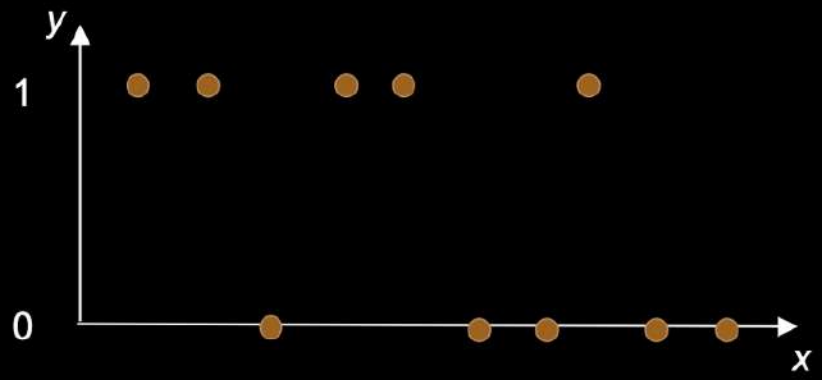
(i) 0% because each of the 10 test examples is classified incorrectly because the majority class of the other 9 is in the opposite class, so the average accuracy is 0%

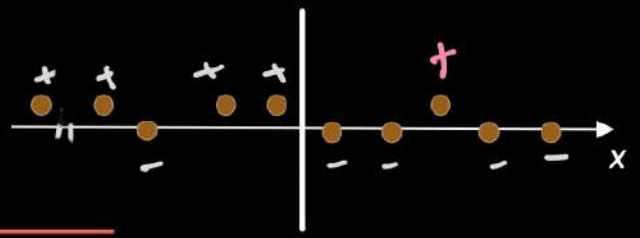
Question:

(c) [3] What is the *2-fold Cross-Validation accuracy*? Assume the leftmost 5 points (i.e., the 5 points with smallest x values) are in one fold and the rightmost 5 points are in the second fold.

(i) 0%
(ii) 20%
(iii) 50%
(iv) 80%
(v) 100%

1st pass

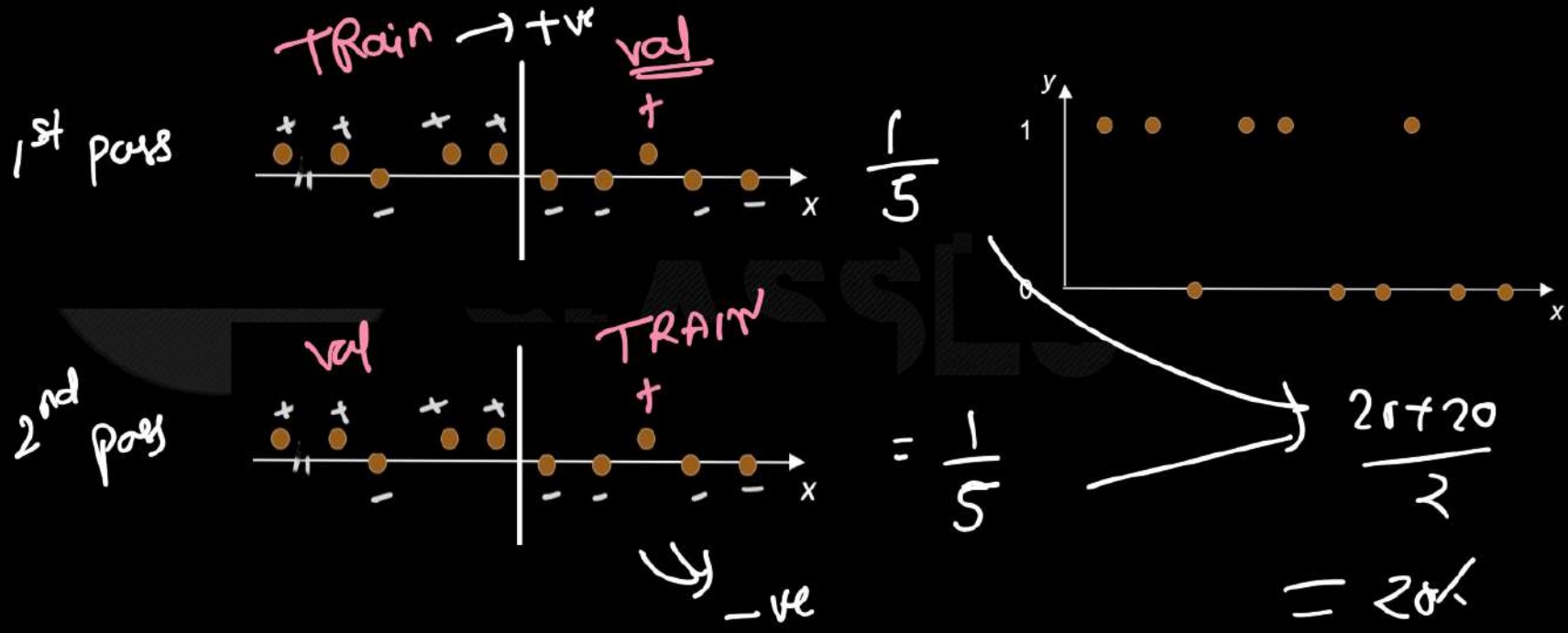




Question:

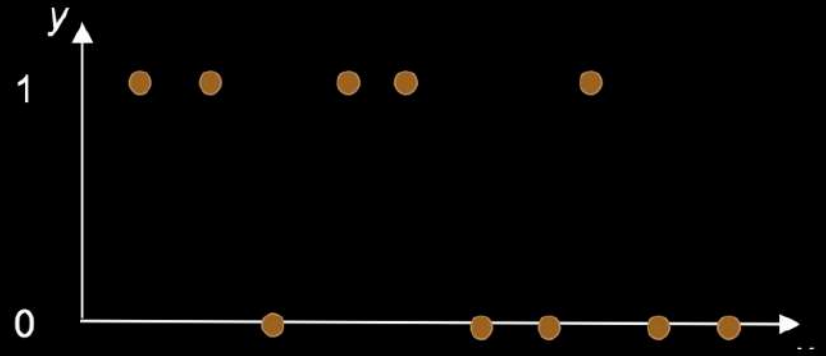
(c) [3] What is the *2-fold Cross-Validation accuracy*? Assume the leftmost 5 points (i.e., the 5 points with smallest x values) are in one fold and the rightmost 5 points are in the second fold.

(i) 0%
(ii) 20%
(iii) 50%
(iv) 80%
(v) 100%



(c) [3] What is the *2-fold Cross-Validation accuracy*? Assume the leftmost 5 points (i.e., the 5 points with smallest x values) are in one fold and the rightmost 5 points are in the second fold.

(i) 0%
(ii) 20%
(iii) 50%
(iv) 80%
(v) 100%



(ii) 20% because for each fold, only 1 of the 5 test examples is classified correctly, so the average accuracy on the two folds is 20%

Question:

Suppose we have a training dataset of 90 points, and a test set of 30 points, and want to know which λ value is best for a ridge regression model. Our candidate hyperparameters are $\lambda = 0.1$, $\lambda = 1$, and $\lambda = 10$.

Suppose we select the best choice of λ from the three choices available using **3-fold** cross validation. As mentioned in class, we can compute the optimal parameters for a ridge regression model with the expression $\vec{\beta} = (\mathbb{X}^T\mathbb{X} + n\lambda I)^{-1}\mathbb{X}^T\vec{y}$. Assume that we use this closed equation to fit the parameters for our model.

i. [2 Pts] During the entire process of selecting our best λ, how many total times will we evaluate the expression $(\mathbb{X}^T\mathbb{X} + n\lambda I)^{-1}\mathbb{X}^T\vec{y}$?

◯ 1 ◯ 2 ◯ 3 ◯ 6 ◯ 9 ◯ 30 ◯ 60 ◯ 90 ◯ 270

ii. [2 Pts] How many rows will be in $\mathbb{X}$ each time this expression is evaluated?

◯ 1 ◯ 2 ◯ 3 ◯ 6 ◯ 9 ◯ 30 ◯ 60 ◯ 90 ◯ 120
◯ It will vary each time. ◯ Not enough information.

Question:

Suppose we have a training dataset of 90 points, and a test set of 30 points, and want to know which λ value is best for a ridge regression model. Our candidate hyperparameters are $\lambda = 0.1$, $\lambda = 1$, and $\lambda = 10$.

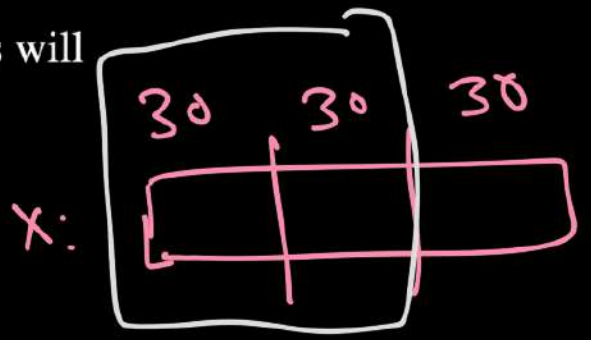
Suppose we select the best choice of λ from the three choices available using **3-fold** cross validation. As mentioned in class, we can compute the optimal parameters for a ridge regression model with the expression $\vec{\beta} = (\mathbb{X}^T\mathbb{X} + n\lambda I)^{-1}\mathbb{X}^T\vec{y}$. Assume that we use this closed equation to fit the parameters for our model.

i. [2 Pts] During the entire process of selecting our best λ, how many total times will we evaluate the expression $(\mathbb{X}^T\mathbb{X} + n\lambda I)^{-1}\mathbb{X}^T\vec{y}$?

○ 1 ○ 2 ○ 3 ○ 6 ○ 9 ○ 30 ○ 60 ○ 90 ○ 270

ii. [2 Pts] How many rows will be in $\mathbb{X}$ each time this expression is evaluated?

○ 1 ○ 2 ○ 3 ○ 6 ○ 9 ○ 30 ○ 60 ○ 90 ○ 120
○ It will vary each time. ○ Not enough information.



Suppose we select the best choice of λ from the three choices available using **3-fold** cross validation. As mentioned in class, we can compute the optimal parameters for a ridge regression model with the expression $\vec{\beta} = (\mathbb{X}^T\mathbb{X} + n\lambda I)^{-1}\mathbb{X}^T\vec{y}$. Assume that we use this closed equation to fit the parameters for our model.

i. [2 Pts] During the entire process of selecting our best λ, how many total times will we evaluate the expression $(\mathbb{X}^T\mathbb{X} + n\lambda I)^{-1}\mathbb{X}^T\vec{y}$?

◯ 1 ◯ 2 ◯ 3 ◯ 6 ◯ **9** ◯ 30 ◯ 60 ◯ 90 ◯ 270

ii. [2 Pts] How many rows will be in $\mathbb{X}$ each time this expression is evaluated?

◯ 1 ◯ 2 ◯ 3 ◯ 6 ◯ 9 ◯ 30 ◯ **60** ◯ 90 ◯ 120
◯ It will vary each time. ◯ Not enough information.

Question:

Suppose we have a training dataset of 90 points, and a test set of 30 points, and want to know which λ value is best for a ridge regression model. Our candidate hyperparameters are $\lambda = 0.1$, $\lambda = 1$, and $\lambda = 10$.

As in the previous part, suppose we want to select the best λ from the three choices above using **3-fold** cross validation. To evaluate the MSE for a given $\vec{\beta}$, we use the sum of squares: $||\vec{y} - \mathbb{X}\vec{\beta}||_2^2$. Reminder that this expression is just another way of writing $\sum(\vec{y}_i - \vec{x}_i^T\vec{\beta})^2$.

i. [2 Pts] During the entire process of selecting our best λ, how many times will this expression get evaluated?

◯ 1 ◯ 2 ◯ 3 ◯ 6 ◯ 9 ◯ 30 ◯ 60 ◯ 90

ii. [2 Pts] How many rows will be in $\mathbb{X}$ each time this expression is evaluated?

◯ 1 ◯ 2 ◯ 3 ◯ 6 ◯ 9 ◯ 30 ◯ 60 ◯ 90 ◯ 120
◯ It will vary each time. ◯ Not enough information.

MSE

we evaluate MSE 9 times on Val. set

https://ds100.org/fa20/resources/assets/exams/fa19/fa19midterm2sol.pdf

Question:

Suppose we have a training dataset of 90 points, and a test set of 30 points, and want to know which λ value is best for a ridge regression model. Our candidate hyperparameters are $\lambda = 0.1$, $\lambda = 1$, and $\lambda = 10$.

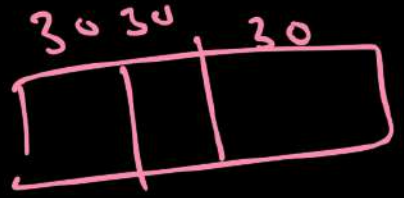
As in the previous part, suppose we want to select the best λ from the three choices above using **3-fold** cross validation. To evaluate the MSE for a given $\vec{\beta}$, we use the sum of squares: $||\vec{y} - \mathbb{X}\vec{\beta}||_2^2$. Reminder that this expression is just another way of writing $\sum(\vec{y}_i - \vec{x}_i^T\vec{\beta})^2$.

i. [2 Pts] During the entire process of selecting our best λ, how many times will this expression get evaluated?

○ 1 ○ 2 ○ 3 ○ 6 ○ 9 ○ 30 ○ 60 ○ 90

ii. [2 Pts] How many rows will be in $\mathbb{X}$ each time this expression is evaluated?

○ 1 ○ 2 ○ 3 ○ 6 ○ 9 ○ 30 ○ 60 ○ 90 ○ 120
○ It will vary each time. ○ Not enough information.



MSE

X :

we evaluate MSE 9 times on Val. set

MSE calculation for val set

https://ds100.org/fa20/resources/assets/exams/fa19/fa19midterm2sol.pdf

Suppose we have a training dataset of 90 points, and a test set of 30 points, and want to know which λ value is best for a ridge regression model. Our candidate hyperparameters are $\lambda = 0.1$, $\lambda = 1$, and $\lambda = 10$.

As in the previous part, suppose we want to select the best λ from the three choices above using **3-fold** cross validation. To evaluate the MSE for a given $\vec{\beta}$, we use the sum of squares: $||\vec{y} - \mathbb{X}\vec{\beta}||_2^2$. Reminder that this expression is just another way of writing $\sum(\vec{y}_i - \vec{x}_i^T\vec{\beta})^2$.

i. [2 Pts] During the entire process of selecting our best λ, how many times will this expression get evaluated?

◯ 1 ◯ 2 ◯ 3 ◯ 6 ◯ **9** ◯ 30 ◯ 60 ◯ 90

ii. [2 Pts] How many rows will be in $\mathbb{X}$ each time this expression is evaluated?

◯ 1 ◯ 2 ◯ 3 ◯ 6 ◯ 9 ◯ **30** ◯ 60 ◯ 90 ◯ 120
◯ It will vary each time. ◯ Not enough information.

https://ds100.org/fa20/resources/assets/exams/fa19/fa19midterm2sol.pdf

Question: True/False

(a) **(1.0 pt)** The test set is divided into k folds. For each fold of the test set, we use the entire training set to train the model, and use the given fold/subset of the test set for validation. The average error among all k folds is the cross-validation error.

True or False: This modification will result in overfitting.

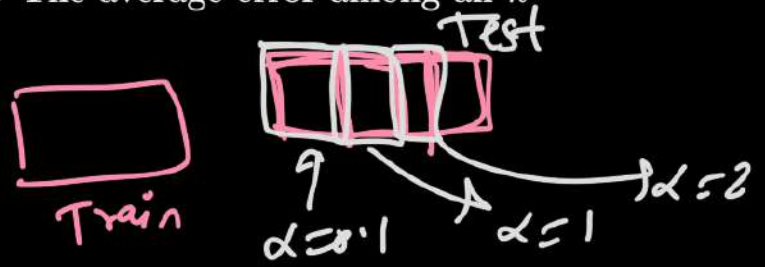
- ◯ True
- ◯ False

(a) **(1.0 pt)** The test set is divided into k folds. For each fold of the test set, we use the entire training set to train the model, and use the given fold/subset of the test set for validation. The average error among all k folds is the cross-validation error.

True or False: This modification will result in overfitting.

- (●) True ✓
- ( ) False

We shouldn't be using the test set for validation purposes; that defeats the purpose of cross-validation.



α = 1 final

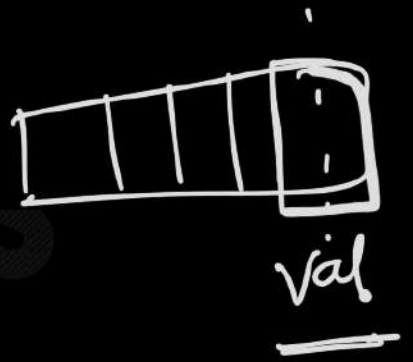
https://ds100.org/fa20/resources/assets/exams/fa20/fa20finalsol.pdf

Question: True/False

(b) (1.0 pt) We use normal k-fold cross-validation, but for each fold we only use half of the validation set for validation.

True or False: This modification will result in overfitting.

- ◯ True
- ◯ False



https://ds100.org/fa20/resources/assets/exams/fa20/fa20finalsol.pdf

(b) (1.0 pt) We use normal k-fold cross-validation, but for each fold we only use half of the validation set for validation.

True or False: This modification will result in overfitting.

○ True

● False

This will not cause overfitting, but it is essentially throwing away data; we could be training our model on more data without overfitting.

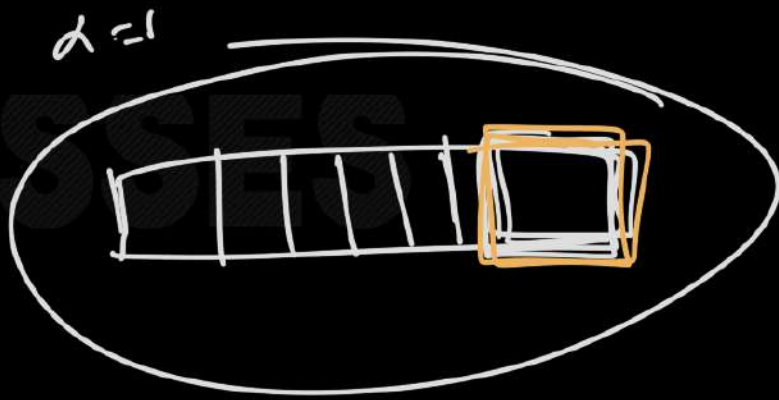
https://ds100.org/fa20/resources/assets/exams/fa20/fa20finalsol.pdf

Question: True/False

(c) (1.0 pt) We use normal k-fold cross-validation, but for each fold we use the entire training set for training.

True or False: This modification will result in overfitting.

- ◯ True
- ◯ False



https://ds100.org/fa20/resources/assets/exams/fa20/fa20finalsol.pdf

(c) (1.0 pt) We use normal k-fold cross-validation, but for each fold we use the entire training set for training.

True or False: This modification will result in overfitting.

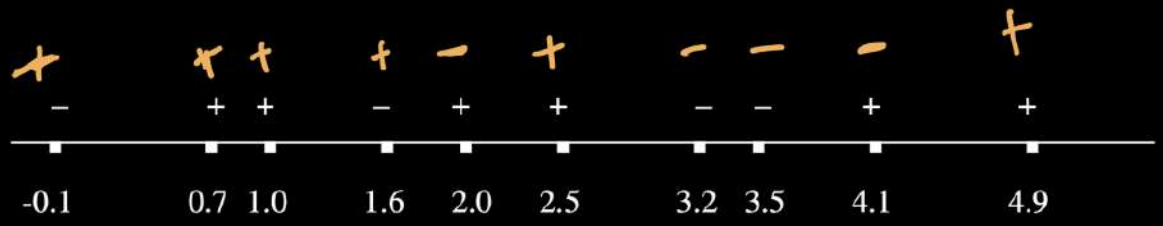
- ● True
- ○ False

The purpose of training on $k-1$ folds and using the remaining fold for validation is to not train and validate our model on the same fold. By making the modification proposed in the question, we would be doing just that.

https://ds100.org/fa20/resources/assets/exams/fa20/fa20finalsol.pdf

Question:

Consider the following dataset with one real-valued input x and one binary output y. We are going to use k-NN with unweighted Euclidean distance to predict y for x.



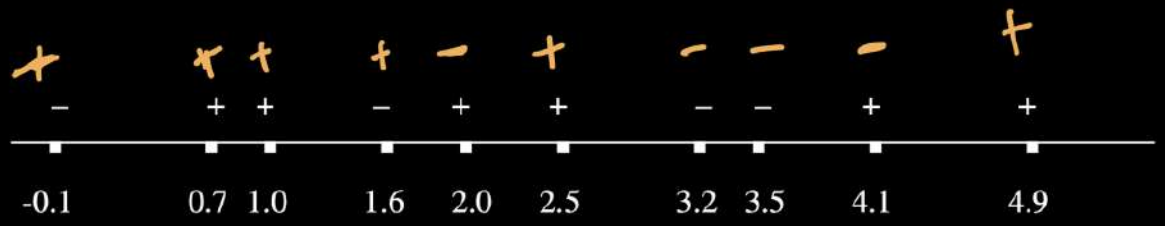
X	Y
-0.1	-
0.7	+
1.0	+
1.6	-
2.0	+
2.5	+
3.2	-
3.5	-
4.1	+
4.9	+

(a) What is the leave-one-out cross-validation error of 1-NN on this dataset? Give your answer as the number of misclassifications.

(b) What is the leave-one-out cross-validation error of 3-NN on this dataset? Give your answer as the number of misclassifications.

https://www.cs.cmu.edu/~epxing/Class/10701/exams/final2003-solution.pdf

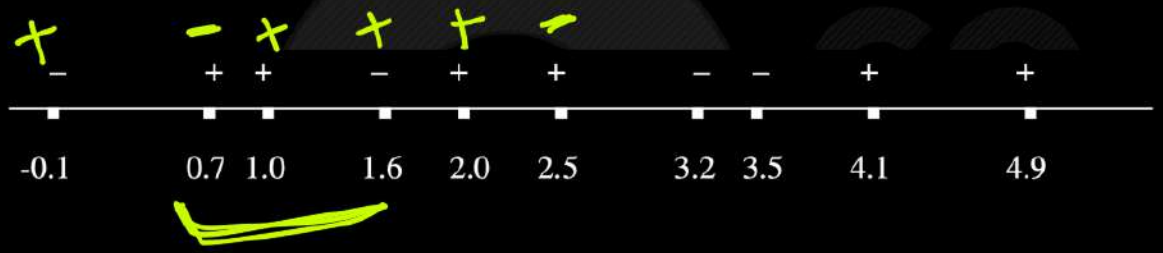
Consider the following dataset with one real-valued input x and one binary output y. We are going to use k-NN with unweighted Euclidean distance to predict y for x.



X	Y
-0.1	-
0.7	+
1.0	+
1.6	-
2.0	+
2.5	+
3.2	-
3.5	-
4.1	+
4.9	+

(a) What is the leave-one-out cross-validation error of 1-NN on this dataset? Give your answer as the number of misclassifications.

Consider the following dataset with one real-valued input x and one binary output y. We are going to use k-NN with unweighted Euclidean distance to predict y for x.



X	Y
-0.1	-
0.7	+
1.0	+
1.6	-
2.0	+
2.5	+
3.2	-
3.5	-
4.1	+
4.9	+

(b) What is the leave-one-out cross-validation error of 3-NN on this dataset? Give your answer as the number of misclassifications.

Consider the following dataset with one real-valued input x and one binary output y. We are going to use k-NN with unweighted Euclidean distance to predict y for x.

X	Y
-0.1	-
0.7	+
1.0	+
1.6	-
2.0	+
2.5	+
3.2	-
3.5	-
4.1	+
4.9	+

(a) What is the leave-one-out cross-validation error of 1-NN on this dataset? Give your answer as the number of misclassifications.

4

(b) What is the leave-one-out cross-validation error of 3-NN on this dataset? Give your answer as the number of misclassifications.

8

MSQ Question:

Mark all the statements that are true, about cross-validation.

- ☐ Increasing k in k-fold cross validation decreases the bias of the model.
- ☐ We use cross validation because it is less computationally intensive than regular validation.
- ☐ We use cross validation over regular validation because cross validation allows us to use the entire training data set to test the model
- ☐ Cross-validation cannot be used in production, as the model is cheating by looking at validation samples.

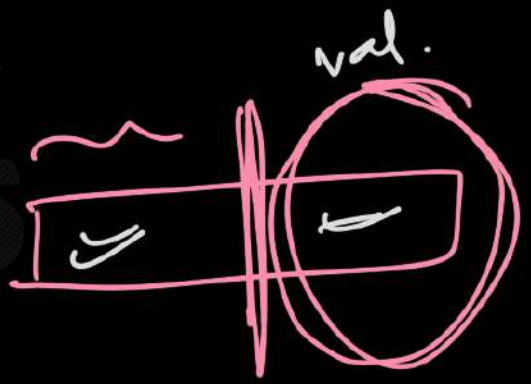
MSQ Question:

Mark all the statements that are true, about cross-validation.

- ☐ Increasing k in k-fold cross validation decreases the bias of the model.
- ☐ We use cross validation because it is less computationally intensive than regular validation.
- ☐ We use cross validation over regular validation because cross validation allows us to use the entire training data set to test the model
- ☐ Cross-validation cannot be used in production, as the model is cheating by looking at validation samples.

Mark all the statements that are true, about cross-validation.

- ☐ **Increasing k in k-fold cross validation decreases the bias of the model.**
- ☐ We use cross validation because it is less computationally intensive than regular validation.
- ☐ **We use cross validation over regular validation because cross validation allows us to use the entire training data set to test the model**
- ☐ Cross-validation cannot be used in production, as the model is cheating by looking at validation samples.

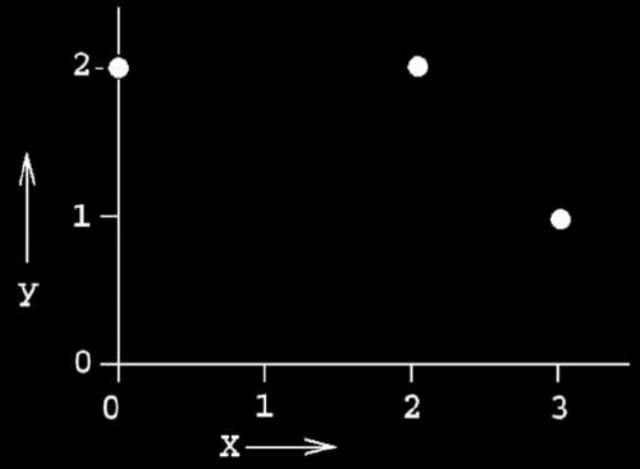


Solution:

- True. Increasing k increases the size of the training set per fold and decreases the size of the validation set. Each model is trained on more data; therefore decreasing the bias of the model.
- False. Cross validation is more computationally intensive than regular validation.
- True.
- False.

Question:

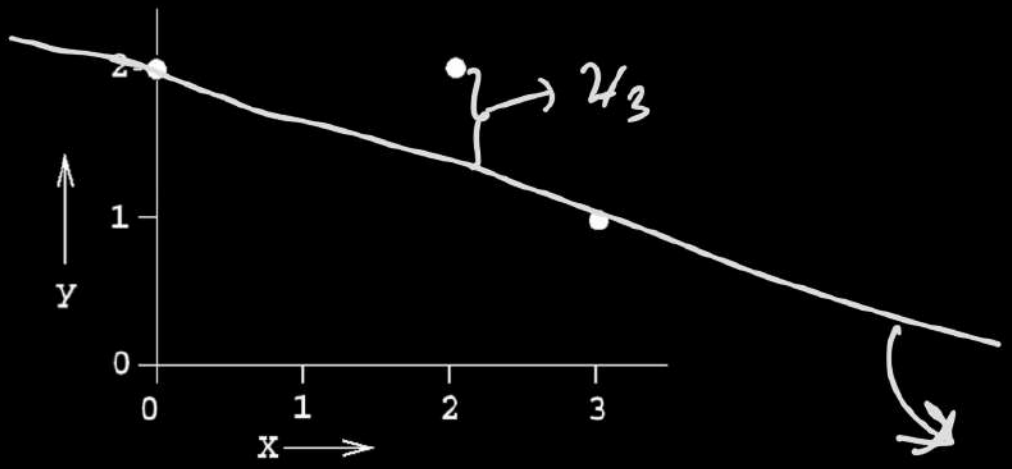
(c) [*5 pts*] Suppose you have this data set with one real-valued input and one real-valued output:



x	y
0	2
2	2
3	1

(c.1) What is the mean squared leave one out cross validation error of using linear regression ? (i.e. the mode is $y = \beta_0 + \beta_1 x$)

(c) [5 *pts*] Suppose you have this data set with one real-valued input and one real-valued output:



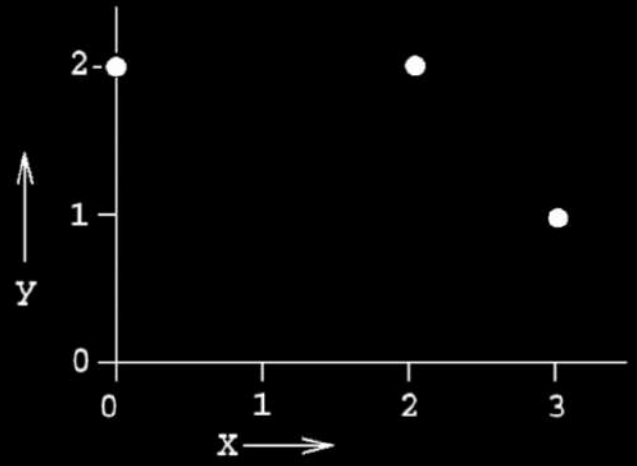
x	y
0	2
2	2
3	1

(c.1) What is the mean squared leave one out cross validation error of using linear regression ? (i.e. the mode is $y = \beta_0 + \beta_1 x$)

Answer: $\frac{2^2+(2/3)^2+1^2}{3} = 49/27$

Question:

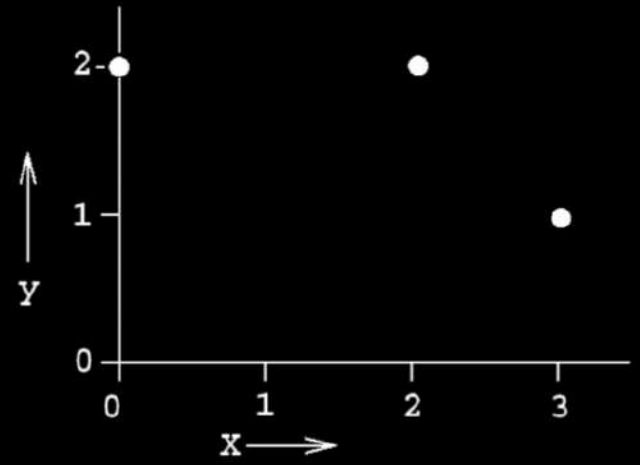
(c) [5 *pts*] Suppose you have this data set with one real-valued input and one real-valued output:



x	y
0	2
2	2
3	1

(c.2) Suppose we use a trivial algorithm of predicting a constant $y = c$. What is the mean squared leave one out error in this case? (Assume c is learned from the non-left-out data points.)

(c) [*5 pts*] Suppose you have this data set with one real-valued input and one real-valued output:



x	y
0	2
2	2
3	1

(c.2) Suppose we use a trivial algorithm of predicting a constant $y = c$. What is the mean squared leave one out error in this case? (Assume c is learned from the non-left-out data points.)

Answer: $\frac{0.5^2+0.5^2+1^2}{3} = 1/2$

Question:

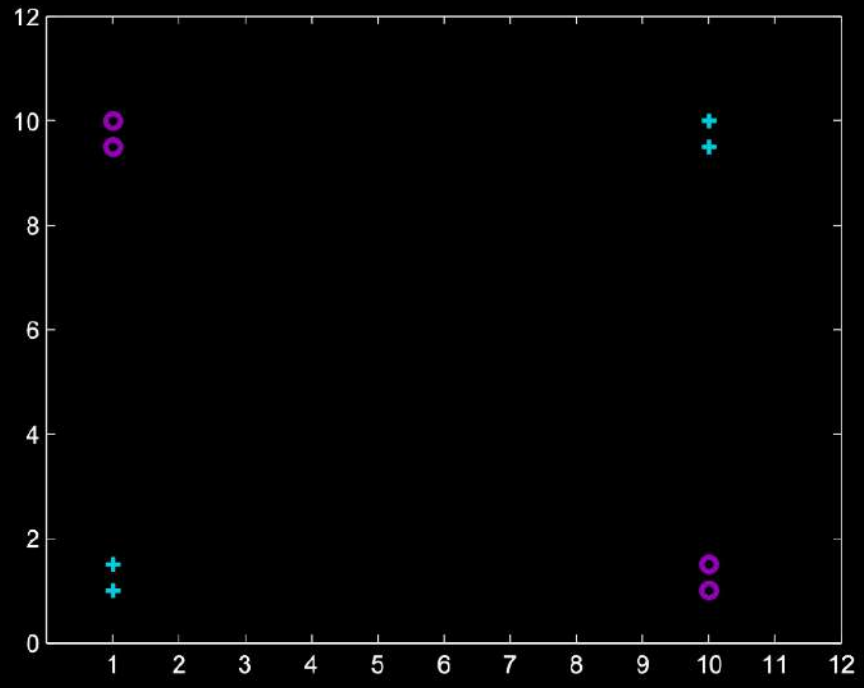
Given the 2D dataset, which of the following is true regarding the performance of 1-nearest neighbor (1-NN) and Support Vector Machines (SVM) in terms of leave-one-out cross-validation error (LOO error)?

A) 1-nearest neighbor (1-NN) has lower LOO error than SVM.

B) SVM has lower LOO error than 1-NN.

C) Both 1-NN and SVM have the same LOO error.

D) It's impossible to compare the LOO error between 1-NN and SVM for this dataset.



Question:

Given the 2D dataset, which of the following is true regarding the performance of 1-nearest neighbor (1-NN) and Support Vector Machines (SVM) in terms of leave-one-out cross-validation error (LOO error)?

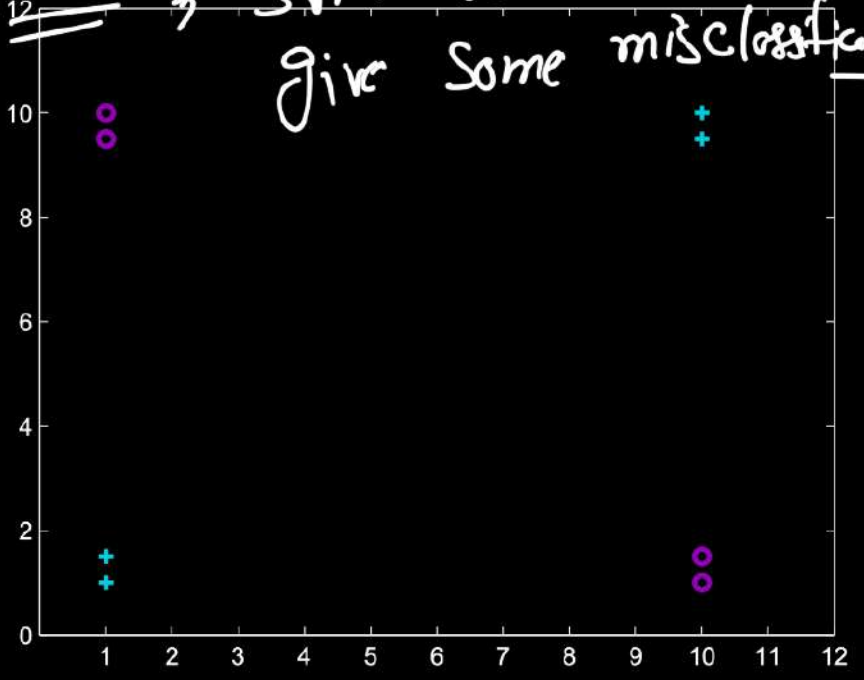
→ 0 misclassifications using 1-NN → SVM will always give some misclassifica

A) 1-nearest neighbor (1-NN) has lower LOO error than SVM.

B) SVM has lower LOO error than 1-NN.

C) Both 1-NN and SVM have the same LOO error.

D) It's impossible to compare the LOO error between 1-NN and SVM for this dataset.



Question:

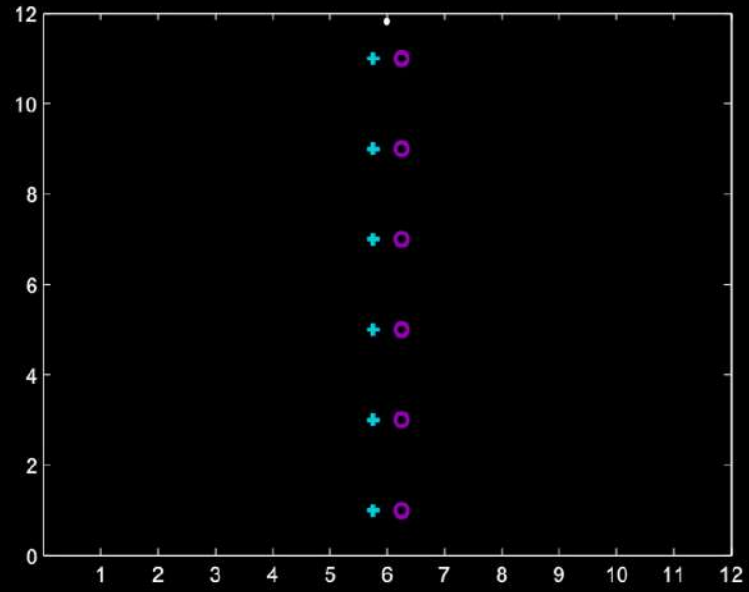
Given the 2D dataset, which of the following is true regarding the performance of 1-nearest neighbor (1-NN) and Support Vector Machines (SVM) in terms of leave-one-out cross-validation error (LOO error)?

A) 1-nearest neighbor (1-NN) has lower LOO error than SVM.

B) SVM has lower LOO error than 1-NN.

C) Both 1-NN and SVM have the same LOO error.

D) It's impossible to compare the LOO error between 1-NN and SVM for this dataset.



Question:

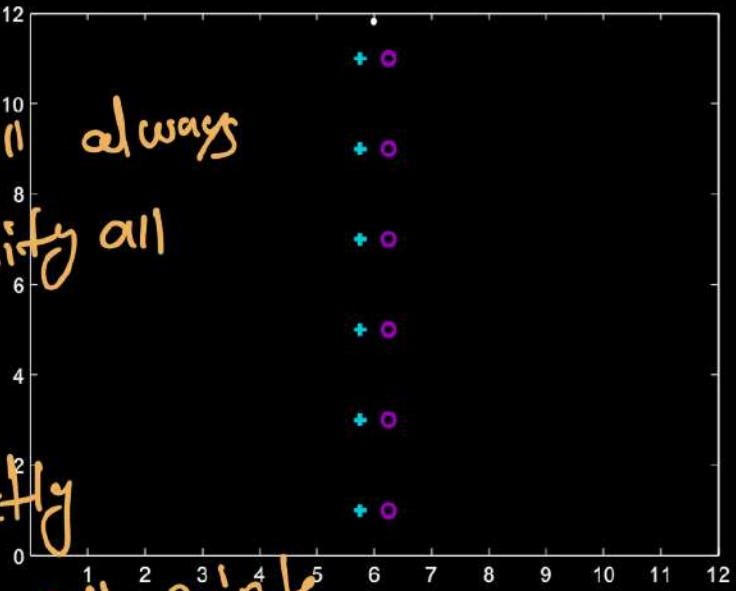
Given the 2D dataset, which of the following is true regarding the performance of 1-nearest neighbor (1-NN) and Support Vector Machines (SVM) in terms of leave-one-out cross-validation error (LOO error)?

A) 1-nearest neighbor (1-NN) has lower LOO error than SVM.

B) SVM has lower LOO error than 1-NN.

C) Both 1-NN and SVM have the same LOO error.

D) It's impossible to compare the LOO error between 1-NN and SVM for this dataset.



1NN will always misclassify all points

SVM correctly classify all points

GO CLASSES

Question:

Consider a scenario where we use leave-one-out cross-validation (LOOCV) with Support Vector Machines (SVM) for binary classification. Given a dataset $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\}$, the cross-validation error for a sample $(\mathbf{x}_n, y_n)$ is defined as:

$$e_n = \begin{cases} 1 & \text{if } \hat{y}_n \neq y_n \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{y}_n$ is the predicted label for $\mathbf{x}_n$ based on the model trained without the sample $\mathbf{x}_n$. The overall cross-validation error is given by:

$$E_{\text{CV}} = \frac{1}{N} \sum_{n=1}^{N} e_n$$

Which of the following statements is true regarding the relationship between the overall cross-validation error E_{CV} and the number of support vectors K?

A. $E_{\text{CV}} \geq \frac{K}{N}$

B. $E_{\text{CV}} = \frac{K}{N}$

C. $E_{\text{CV}} \leq \frac{K}{N}$

D. E_{CV} is independent of K

https://exams.skule.ca/exams/ECE421H1_20201_631619146406Final_Winter_2020.pdf

GO CLASSES

Question:

Consider a scenario where we use leave-one-out cross-validation (LOOCV) with Support Vector Machines (SVM) for binary classification. Given a dataset $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\}$, the cross-validation error for a sample $(\mathbf{x}_n, y_n)$ is defined as:

$$e_n = \begin{cases} 1 & \text{if } \hat{y}_n \neq y_n \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{y}_n$ is the predicted label for $\mathbf{x}_n$ based on the model trained without the sample $\mathbf{x}_n$. The overall cross-validation error is given by:

$$E_{\text{CV}} = \frac{1}{N} \sum_{n=1}^{N} e_n$$

→ No. of misclas.

Which of the following statements is true regarding the relationship between the overall cross-validation error E_{CV} and the number of support vectors K?

A. $E_{\text{CV}} \geq \frac{K}{N}$

B. $E_{\text{CV}} = \frac{K}{N}$

C. $E_{\text{CV}} \leq \frac{K}{N}$

D. E_{CV} is independent of K

$E_{CV} \leq \frac{K}{N}$

V G Masilamani to Everyone 9:58 AM
VG
c

Dev to Everyone 9:59 AM
D
C

044_sumaya to Everyone 10:00 AM

C

https://exams.skule.ca/exams/ECE421H1_20201_631619146406Final_Winter_2020.pdf

GO CLASSES

50 questions in CV

→ Precision, Recall, F1 score

interesting & less understood.