Dictionary-based Representations Pick 3 questions, 1 pts per question

Image Denoising

Which of the following statements is False about image denoising?

Image denoising is one application of K-SVD.

Corrupted images cannot be used to learn a dictionary

Methods like Matching Pursuit or Basis Pursuit can be useful in image denoising.

Correct Answer

2

All of them

1

3

1, 3

SVD K-means

It was covered in class that methods like K-means, SVD K-means can be useful to build a dictionary. Which of the following statements are reasonable statements about these methods?

Dictionary atoms you obtained using K-means is simply the cluster means of input data.

Given a dictionary obtained using length-unconstrained K-means clustering, the weight W to be learned can only be 1 sparse, and ||W||1 must be 1.

In SVD K-means method, to obtain dictionary atoms, we need to find the principal eigenvector for each cluster

Correct Answer

1, 3

2, 3

1, 2

1, 2, 3

Basis Pursuit

Which of the following statements are true about matching pursuit and basis pursuit?

Matching pursuit does greedy optimization at each step.

Matching pursuit uses the minimum inner product to find best match basis for updated signal.

Basis pursuit generally takes more time than matching pursuit.

Increasing penalty term λ in basis pursuit will promote sparsity.

2, 3

Correct Answer

1, 4

1, 3, 4

1, 2, 4

k-means

Which of the following statements is true?

The k-means algorithm for clustering (with L2 error as an objective function) is guaranteed to converge to a local optimum.

Given a fixed number of clusters k, globally minimizing the k-means objective function is NP-hard.

The k-means algorithm does coordinate descent on a non-convex objective function.

Correct Answer

All of them

Only 1

Only 2

2 and 3

1 and 3

1 and 2

Overcomplete Representations

Which of the following statements are true about overcomplete representations?

The number of atoms in an overcomplete dictionary D is larger than dimensions of the input data.

Given input data and overcomplete dictionary, the weight to be learned is unique.

Dictionary-based overcomplete representations are semantically more interpretable than component analysis (e.g. PCA, ICA).

Correct Answer

1, 3

All of them

2, 3

1, 2

K-SVD

Which of the following statements are true about K-SVD?

The dictionary obtained by performing K-SVD is a dictionary where any data vector can be composed of at least K dictionary entries.

Although finding an optimal dictionary for a given dataset is a non-convex problem, K-SVD can still guarantee to find global optimum by doing an iterative update.

K-SVD algorithm is not computationally demanding at all.

In each updating iteration (shown in the slides), we replaced the kth atom with the principal eigenvector of {ek(x)}

1, 2, 4

All of them

Correct Answer

only 4

2, 4

Linear Classifiers Pick 4 questions, 1 pts per question

SVM

In a soft-margin support vector machine, decreasing the slack penalty term, C causes:

More overfitting

A smaller margin

Less overfitting

Less sensitivity to outliers

Correct Answer

3 and 4

1 and 3

1 and 2

all of them

Linear SVM

Let S1 and S2 be the set of support vectors and w1 and w2 be the learnt weight vectors for a linearly separable problem using hard and soft margin linear SVMs respectively. Which of the followings are correct?

S1 ⊂ S2

w1 = w2

S1 may not be a subset of S2

w1 may not be equal to w2

1 and 2

1 and 4

Correct Answer

3 and 4

2 and 3

SVM Classifier

Which of the following can help to reduce overfitting in an SVM classifier?

Normalizing the data

Correct Answer

Use of slack variables

High-degree polynomial features

Setting a very low learning rate

SVM

Which of the following is true?

In One vs All Multi-Class Classification in SVM, we are trying to classify an input data point X as one of the N classes (C1,..., Cn), each of which has a parameter vector w1,...,wn. We classify point X as the class Ci which maximizes the inner product of X and wi

SVMs directly give us the posterior probabilities P(y = 1| x) and P(y = 1| x).

In SVMs, the sum of the Lagrange multipliers corresponding to the positive examples is equal to the sum of the Lagrange multipliers corresponding to the negative examples.

1 and 2

only 3

only 1

Correct Answer

1 and 3

SVM

Which of the following statements about Support Vector Machine is FALSE?

By using Kernel Trick, data points can be mapped to higher dimensional space to make them linearly separable

We need to minimize the distance between margin planes while minimizing total slack distance, so two objectives must be a trade-off

Non-linear separable classification problem can be solved by SVM sometimes.

3

1, 2, 3

Correct Answer

2

1, 3

Hyperplane

The shortest distance from a point z to a hyperplane wTx = 0 is

Correct Answer

LaTeX: \frac{w^\top z}{\|w \|}w ⊤ z ‖ w ‖

LaTeX: w^\top zw ⊤ z

LaTeX: \frac{w^\top z}{\|w \|^2}w ⊤ z ‖ w ‖ 2

LaTeX: \|w \| \cdot \|z\|‖ w ‖ ⋅ ‖ z ‖

Hard-margin SVM

Suppose we train a hard-margin linear SVM on n >100 data points in LaTeX: \mathbb{R}^2R 2, yielding a hyperplane with exactly 2 support vectors. If we add one more data point and retrain the classifier, what is the maximum possible number of support vectors for the new hyperplane (assuming the n + 1 points are linearly separable)?

n

2

Correct Answer

n + 1

3

SVM

If we remove a support vector from an SVM, the size of the margin could:

Decrease

Stay the same

Increase

Correct Answer

2 and 3

only 1

only 3

1 and 2

Regression Pick 3 questions, 1 pts per question

Regularization 1

Which of the following is true of regularization?

L1 regularization can be differentiable

L1 regularization can be used for feature selection in ML problems

L2 regularization results in a sparse solution because it discards features that contributed nothing to the solution

Correct Answer

only 2

1 and 2

only 3

2 and 3

All of them

1 and 3

Regularization 2

Which of the following is true of regularization?

The general idea of regularization is to ensure that a function generalizes well on test data only

If you have a problem involving multiple simultaneous objectives, you can also rewrite it in terms of Lagrangian

Too strong regularization penalty will overfit the training data while too weak penalty will underfit the problem

1 and 2

2, 3

All of them

None

Correct Answer

only 2

linear regression

Which of the following statements are true about linear regression?

Linear regression cannot be done in high dimensional space.

Linear regression can be useful in trend forecasting.

By doing linear regression, we can know how much the thing we want to estimate depends on each observable input variable.

1, 2

1, 2, 3

Correct Answer

2, 3

1, 3

Nearest-neighbor

Which of the following options is true about Nearest-neighbor based prediction?

It predicts the weighted average value of desired attribute from all the training instances.

It can be used in both classification and regression.

The weight of each neighbor has an inverse relationship to distance from the neighbor to the point

Normalizing axis is needed because the weights of the training instances will change if any axis is stretched.

Correct Answer

All of them

1, 2, 4

1, 2, 3

2, 3, 4

False Statement

Which of the following statements are false?

For a given x the linear regression prediction is a weighted combination of values in training data

Selecting smaller values of k in k-nearest neighbours provides smoother fit of the data.

Selecting a big value of k would eventually assign the same label to all points.

Selecting a small value of k does not lead to overfitting.

2 and 3

1 and 2

Correct Answer

2 and 4

1 and 3

k-nearest neighbour

Which of the following statements is/are true about regression and k-nearest neighbour

Regularization helps reduce the effect of outliers on the weights

The dimensionality of data does not affect the prediction accuracy (on test data) of linear regression

The dimensionality of data does not affect the prediction accuracy (on test data) of k-nearest neighbour

The same closed-form solution can be obtained for least square regression in both cases when the dimensionality is greater than the numbers of observations or vice versa.

only 1

Correct Answer

1 and 2

1 and 3

1, 3, 4