NMF and Clustering Pick 10 questions, 1 pts per question

Sparsity Constraint

To impose sparsity in modified NMF model (shown in the slides), which of the following options apply?

1. Combine divergence with LaTeX: \ell\_0ℓ 0 norm of LaTeX: WW (weight matrix) using combinatorial optimization techniques

2. Combine divergence with LaTeX: \ell\_1ℓ 1 norm of LaTeX: WW

3. Combine divergence with LaTeX: \ell\_2ℓ 2 norm of LaTeX: WW

4. Minimize LaTeX: \ell\_1ℓ 1 norm can be equivalent to minimizing LaTeX: \ell\_0ℓ 0 norm.

Correct Answer

3 is not true

2 and 4 are true

All of the above are true

1 and 2 are not always true

NMF Trivia

Which of the following statements are true about Non-negative Matrix Factorization?

1. NMF can be used to predict missing data

2. Non-negativity can result in semantically meaningful bases

3. NMF can be employed to do real time polyphonic pitch tracking

4. NMF's positive coefficients imply an additive combination of basis

Correct Answer

All of the above

1, 2, 4

1, 2, 3

1 and 2

NMF Optimization

In NMF we are solving the optimization problem LaTeX: \min\_{B, W}||V - BW||^2min B , W | | V − B W | | 2, where LaTeX: V \in R^{m \times n}V ∈ R m × n, LaTeX: B \in R\_+^{m \times k }B ∈ R + m × k and LaTeX: W \in R^{k \times n }\_+W ∈ R + k × n. Which of the following is true about this optimization problem?

1. This is a convex optimization problem

2. For a fixed LaTeX: BB, it becomes a convex optimization problem

3. For a fixed LaTeX: WW, it becomes a convex optimization problem

4. First order methods like gradient descent cannot be used for solving NMF

Correct Answer

2 and 3

1, 2, and 3

2, 3, and 4

1 and 4

NMF and PCA

NMF and PCA can both be used for dimensionality reduction. Which of the following is true about these methods?

1. If the data matrix LaTeX: XX is given to be non-negative then PCA and NMF will give same solution.

2. Both PCA and NMF make the assumption that data is multivariate Gaussian.

3. If the number of bases from NMF or PCA is equal to dimension of data in LaTeX: XX, then we can have bases which will give perfect reconstruction.

4. Both NMF and PCA learns orthogonal bases.

Correct Answer

Only 3

2 and 3

1 and 3

3 and 4

Sparsity Trivia

Which of the following statements are true about NMF?

1. In LaTeX: V=BWV = B W, if LaTeX: BB has fewer columns than rows, it is called an overcomplete representation.

2. Sparse representations need not be overcomplete, but the reverse will generally not provide useful representations.

3. With sparsity, the data are “pulled towards” the bases.

4. As solutions become more sparse they become more informative, and this happens more often when LaTeX: K \geq DK ≥ D, where LaTeX: KK is the dimensionality of the data and LaTeX: DD are the number of bases.

5. In the 2-D sparsity example from class, if a point is outside the triangle provided by the bases, we can represent it by projecting it onto the triangle with some error.

Correct Answer

2 and 5

All of the above

2, 4, 5

1, 3, 5

1, 2, 4

K-means Basics

Suppose we are given data comprising points of several different classes. Each class has a different probability distribution from which the sample points are drawn. We do not have the class labels. We use k-means clustering to try to guess the classes. Which of the following circumstances would undermine its effectiveness?

1. Some of the classes are not normally distributed

2. Each class has the same mean

3. The variance of each distribution is small in all directions

4. You choose k = n, the number of sample points

Correct Answer

2 and 4

1 and 3

Only 4

All of the above

K-means Convergence

Which of the following statements is true?

1. The k-means algorithm for clustering is guaranteed to converge to a local optimum.

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

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

Correct Answer

All of the above

Only 1

Only 2

2 and 3

K-means Example

Consider the following dataset: LaTeX: A = (0, 2)A = ( 0 , 2 ), LaTeX: B = (0, 1)B = ( 0 , 1 ) and LaTeX: C = (1, 0)C = ( 1 , 0 ). The k-means algorithm is initialized with centers at LaTeX: AA and LaTeX: BB. Upon convergence, the two centers will be at

Correct Answer

LaTeX: AA and the midpoint of LaTeX: BCB C

LaTeX: AA and LaTeX: CC

LaTeX: AA and LaTeX: BB

LaTeX: CC and the midpoint of LaTeX: ABA B

Lloyd's Complexity

Assume Lloyd's algorithm is run on LaTeX: nn samples in LaTeX: \mathbb{R}^dR d and converges after LaTeX: tt iterations. What is its approximate computational complexity?

Correct Answer

LaTeX: O(tKnd)O ( t K n d )

LaTeX: O(t^2Knd)O ( t 2 K n d )

LaTeX: O(tK^2nd)O ( t K 2 n d )

LaTeX: O(tKn^2d)O ( t K n 2 d )

K-means Trivia

Which of the following statements are true about k-means clustering?

1. Different initialization methods will not cause any difference in results

2. K-means clustering is an unsupervised learning method

3. Considering an easy case for k-means, k = 1, the optimal choice of center is LaTeX: c = \sum\_{i=1}^{n}x^{i}c = ∑ i = 1 n x i

Correct Answer

Only 2

All of the above

Only 3

1 and 3

K-means Initialization

Say you want to use k-means clustering to partition N data points, where LaTeX: 1 < k < N1 < k < N. Which of the following is a sensible way to pick the value of k?

1. Choose k such that the k-means cost function is minimized on held out data.

2. Choose k such that LaTeX: k, k+1, ..., Nk , k + 1 , . . . , N have similar values of the cost function but that LaTeX: 1, ..., k-11 , . . . , k − 1 have much higher costs, where all costs are evaluated from training data.

3. Set LaTeX: k = n/2k = n / 2

4. Set LaTeX: k = log(n/2)k = l o g ( n / 2 )

Correct Answer

1, 2

1, 3

2, 3

2, 3, 4

More Kernels

Which of the following statements about kernel functions is/are true?

1. If LaTeX: \Phi:\Re^d \to \Re^kΦ : ℜ d → ℜ k is a function, then LaTeX: k(x,y)=\langle \Phi(x),\Phi(y) \ranglek ( x , y ) = ⟨ Φ ( x ) , Φ ( y ) ⟩ is a valid kernel function. (Here, LaTeX: k>dk > d and both are integers, and LaTeX: \langle a,b \rangle⟨ a , b ⟩ denotes the inner product of LaTeX: aa and LaTeX: bb.)

2. If LaTeX: k\_1k 1 and LaTeX: k\_2k 2 are valid kernel functions, and LaTeX: c\_1c 1 and LaTeX: c\_2c 2 are constants such that LaTeX: c\_1 \geq 0c 1 ≥ 0 and LaTeX: c\_2 \geq 0c 2 ≥ 0, then LaTeX: k(x, y) = c\_1 k\_1(x, y) - c\_2 k\_2(x, y)k ( x , y ) = c 1 k 1 ( x , y ) − c 2 k 2 ( x , y ) is a valid kernel function.

3. Let LaTeX: K(x,y) = \langle \Phi(x),\Phi(y) \rangleK ( x , y ) = ⟨ Φ ( x ) , Φ ( y ) ⟩ be a valid kernel over LaTeX: \mathbb{R}^pR p. Then LaTeX: K\_{norm}(x,y) = \frac{K(x,y)}{\sqrt{K(x,x)K(y,y)}}K n o r m ( x , y ) = K ( x , y ) K ( x , x ) K ( y , y ) is also a valid kernel, i.e. LaTeX: K\_{norm} = \langle \Phi\_{norm}(x),\Phi\_{norm}(y) \rangleK n o r m = ⟨ Φ n o r m ( x ) , Φ n o r m ( y ) ⟩ for some LaTeX: \Phi\_{norm}(x)Φ n o r m ( x ).

Correct Answer

1, 3

1

1, 2

All of the above

Kernels

Which of the following is true about Kernel functions

1. The Kernel matrix LaTeX: KK for a set of vectors LaTeX: \mathbf{x}\_1 \dots \mathbf{x}\_nx 1 … x n has to be positive semi-definite

2. The kernel defined by LaTeX: K(\mathbf{x}\_1, \mathbf{x}\_2) = e^{|\mathbf{x}\_1| - |\mathbf{x}\_2|}K ( x 1 , x 2 ) = e | x 1 | − | x 2 | is a valid kernel

3. The kernel defined by LaTeX: K(\mathbf{x}\_1, \mathbf{x}\_2) = (\mathbf{x}\_1^T\mathbf{x}\_2 + c)^rK ( x 1 , x 2 ) = ( x 1 T x 2 + c ) r, where LaTeX: rr is an integer greater than 2 and LaTeX: c \geq 0c ≥ 0 is a valid kernel

4. If LaTeX: K(\mathbf{x}\_1, \mathbf{x}\_2)K ( x 1 , x 2 ) is a kernel, then it is not necessarily true that LaTeX: K'(\mathbf{x}\_1, \mathbf{x}\_2) = \exp(K(\mathbf{x}\_1, \mathbf{x}\_2))K ′ ( x 1 , x 2 ) = exp ⁡ ( K ( x 1 , x 2 ) ) is also a kernel

Correct Answer

1, 3, 4

1, 3

2, 3, 4

2, 3