

Course > Week 1... > Proble... > Proble...

Problem Set 10

Problems 1-4 correspond to "Autoencoders"

Problem 1

1/1 point (graded)

When we run k-means on a data set of n points in \mathbb{R}^d , and think of it as an autoencoder, what is the number of hidden units?

$\bigcirc n$
$\bigcirc d$
lacksquare
$\log k$
✓
Submit

Problem 2

1/1 point (graded)

Which of the following types of information about a data point x is lost in the k-means autoencoder? Select all that apply.

W	hich cluster x is assigned to		

7/2019	Problem Set 10 Problem Set 10 DSE220x Courseware edX
₩ Where	e x lies relative to its cluster center
The di	istance of x from its cluster center
~	
Submit	
Problem	3
1/1 point (gra Which of th Select all th	e following statements accurately describe the notion of an autoencoder?
✓ It is ar	n abstraction that unifies many different types of unsupervised learning.
lt is ar	n abstraction that unifies many different types of supervised learning.
✓ It abst	tracts the operation of changing the representation of data.
Submit	
Problem	4
	aded) activity of an experimental subject is measured by placing a variety of her head. Each sensor receives a signal that is a linear combination of

electrical activity in different regions of her brain. Based on these sensor readings, we would like to infer the activity in each of these individual brain regions. Which of the following types of unsupervised learning applies most directly to this problem?

Clustering	
Principal component analysis	

Manifold learning
Independent component analysis
✓
Submit
Problems 5-6 correspond to "Distributed representations"
Problem 5
1/1 point (graded) In which of the following ways could \emph{k} -means be used to create a $\emph{distributed}$ representation? Select all that apply.
Encode each point by the closest cluster center.
Encode each point by its distance to the closest cluster center.
Encode each point by its distances to all the cluster centers.
Encode each point by its distance to the furthest cluster center.
✓
Submit

Problem 6

1/1 point (graded)

One compromise between a one-hot encoding and a dense distributed encoding is a *sparse distributed encoding*. Here, the hidden representation of an input is a sparse vector, in which only a small number of the entries (say, ℓ of them) are non-zero. Which of the following is a sensible way of using k-means to produce a representation of this type?

Encode each	point by its distances to the ℓ closest centers.
Encode each	point by its distances to all the centers.
Encode each	point by its distances to the first ℓ centers.
✓	
Submit	
Problems 7-10 coi	respond to "Feedforward neural networks"
	ural network has five layers, each consisting of 100 nodes, and each the previous layer. Roughly how many parameters does this network
<u> </u>	
10000	
50000	
~	
Submit	
Problem 8	
1/1 point (graded)	

A particular node h in a feedforward neural net has parents z, and sets its own value by computing a linear function of its parents, $w \cdot z + b$, and then applying the rectified linear activation function to the result. For what values of z does h take on a negative value?

igcup When $w\cdot z+b<0$

igcup When $w\cdot z+b>0$

igcup When $w\cdot z+b=0$

 $lackbox{}{{\bf a}}}} \lackbox{}{lackbox{}{{\bf a}}} \lackbox{}{{\bf a}} \la$

~

Submit

Problem 9

1/1 point (graded)

The output layer of a particular neural net has four nodes y_1,y_2,y_3,y_4 , representing four labels in a classification problem. For some input x, these nodes end up with the values

$$y_1 = 2.0, y_2 = 0.0, y_3 = 0.0, y_4 = 1.0,$$

and these are converted to probabilities using a softmax. What is the probability assigned to label 4?

0.08

 \bullet 0.22

0.17

0.61

~

Submit

Problem 10

1/1 point (graded)

It is known that any function over d variables can be arbitrarily well approximated by:

A linear function

igcap A neural net with one hidden layer containing d nodes

 A neural net with one hidden layer containing potentially a large number of nodes

igcap A neural net with depth d, in which each hidden layer has d nodes



Submit

Problems 11-14 correspond to "Training neural networks"

Problem 11

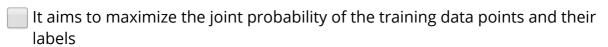
1/1 point (graded)

Which of the following statements are true of the cross-entropy loss function for training a feedforward neural network? Select all that apply.

It is convex

🗸 It potentially has multiple local optima

It aims to maximize the probability of the training data's labels





Submit

Problem 12

1/1 point (graded)

Let $f,g,h:\mathbb{R} o\mathbb{R}$ be some functions, and define function $J:\mathbb{R} o\mathbb{R}$ by $J\left(x
ight)=h\left(g\left(f\left(x
ight)
ight)\right)$. Using the chain rule, what can we say about $J'\left(x
ight)$?

$$igcup J'(x) = h'\left(g'\left(f'\left(x
ight)
ight)
ight)$$

$$\bigcirc J'(x) = h'\left(g\left(f(x)\right)\right)$$

$$\bigcirc J'(x) = h'\left(g\left(f(x)\right)\right)g'\left(f(x)\right)$$

$$\bullet J'\left(x\right) = h'\left(g\left(f\left(x\right)\right)\right)g'\left(f\left(x\right)\right)f'\left(x\right)$$



Answer

Correct: This is two applications of the chain rule.

Submit

Problem 13

1/1 point (graded)

When training a feedforward neural net using stochastic gradient descent, what is involved in a single update?

A forward pass through the net, followed by a backward pass.

A backward pass through the net, followed by a forward pass.

2019	Problem Set 10 Problem Set 10 DSE220x Courseware edX
Just a forwa	rd pass through the net.
Several forw	vard and backward passes through the net.
✓	
Submit	
Problem 14	
1/1 point (graded)	
Which of the follo	wing is an accurate characterization of backpropagation?
	ent way of computing all the derivatives needed for training a net by
gradient des	scent or stochastic gradient descent.
	que for avoiding local optima while training a neural net.
lt is a techni	
It is a techni	que for avoiding local optima while training a neural net.
It is a techni	que for avoiding local optima while training a neural net. native to gradient descent for neural net training.
It is a techni	que for avoiding local optima while training a neural net. native to gradient descent for neural net training.
It is a techni	que for avoiding local optima while training a neural net. native to gradient descent for neural net training.

© All Rights Reserved