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?

- ( ) n
- ( ) d
- (•) k
- ( ) $\log k$

✓

Submit

Answers are displayed within the problem

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.

- [ ] Which cluster x is assigned to
- [x] Where x lies relative to its cluster center
- [x] The distance of x from its cluster center

Submit

Answers are displayed within the problem

Problem 3

1/1 point (graded)
Which of the following statements accurately describe the notion of an autoencoder? Select all that apply.

- [x] It is an abstraction that unifies many different types of unsupervised learning.
- [ ] It is an abstraction that unifies many different types of supervised learning.
- [x] It abstracts the operation of changing the representation of data.

Submit

Answers are displayed within the problem

Problem 4

1/1 point (graded)

The neural activity of an experimental subject is measured by placing a variety of sensors on 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
- (x) Independent component analysis

✔

Submit

Answers are displayed within the problem

Problems 5-6 correspond to "Distributed representations"

Problem 5

1/1 point (graded)

In which of the following ways could k-means be used to create a *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.
- [x] Encode each point by its distances to all the cluster centers.

- [ ] Encode each point by its distance to the furthest cluster center.

Submit

Answers are displayed within the problem

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?

- (x) 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

Answers are displayed within the problem

Problems 7-10 correspond to "Feedforward neural networks"

Problem 7

1/1 point (graded)

A feedforward neural network has five layers, each consisting of 100 nodes, and each fully connected to the previous layer. Roughly how many parameters does this network have?

- ○ 100
- ○ 500
- ○ 10000
- ● 50000

✔

Explanation

The number of edges between any two consecutive layers is $100^2 = 10000$. And there are five such layers.

Submit

Answers are displayed within the problem

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?

- ○ When $w \cdot z + b < 0$
- ○ When $w \cdot z + b > 0$
- ○ When $w \cdot z + b = 0$
- ● h never takes on a negative value

Explanation
The rectified linear activation function zeros out any negative value.

Submit

Answers are displayed within the problem

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
- ◉ 0.22
- ○ 0.17
- ○ 0.61

✔

Hint (1 of 1): The probability assigned to label 4 is $\frac{e^{y_4}}{e^{y_1}+e^{y_2}+e^{y_3}+e^{y_4}}$

Next Hint

Submit

Answers are displayed within the problem

Problem 10

1/1 point (graded)
It is known that any function over d variables can be arbitrarily well approximated by:

- ○ A linear function
- ○ A neural net with one hidden layer containing d nodes
- ● A neural net with one hidden layer containing potentially a large number of nodes
- ○ A neural net with depth d, in which each hidden layer has d nodes

✔

Submit

Answers are displayed within the problem

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

- [x] It aims to maximize the probability of the training data's labels
- [ ] It aims to maximize the joint probability of the training data points and their labels

✓

Submit

Answers are displayed within the problem

Problem 12

1/1 point (graded)

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

- ( ) $J'(x) = h'(g'(f'(x)))$
- ( ) $J'(x) = h'(g(f(x)))$
- ( ) $J'(x) = h'(g(f(x)))\, g'(f(x))$
- (•) $J'(x) = h'(g(f(x)))\, g'(f(x))\, f'(x)$

✓

Answer

Correct: This is two applications of the chain rule.

Submit

Answers are displayed within the problem

Problem 13

1/1 point (graded)

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

- (selected) A forward pass through the net, followed by a backward pass.
- A backward pass through the net, followed by a forward pass.
- Just a forward pass through the net.
- Several forward and backward passes through the net.

✔

Submit

Answers are displayed within the problem

Problem 14

1/1 point (graded)

Which of the following is an accurate characterization of backpropagation?

- (selected) It is an efficient way of computing all the derivatives needed for training a net by gradient descent or stochastic gradient descent.
- It is a technique for avoiding local optima while training a neural net.
- It is an alternative to gradient descent for neural net training.
- It is a randomized scheme for speeding up neural net training.

✔

Submit

Answers are displayed within the problem