BerkeleyX: CS190.1x Scalable Machine Learning

ONE-HOT-ENCODING (1 point possible)

Using one-hot-encoding, a categorical feature with four distinct values would be represented by how many features?

- 1 feature
- 2 features
- 3 features
- 4 features

?

CHECK

RARE EVENTS (1 point possible)

For rare events it is often a good idea to predict probabilities instead of classes.

- True
- False

?

CHECK

FEATURE REPRESENTATION (1 point possible)

The OHE features in the lab are stored in a:

- [ ] Dense representation
- [ ] Sparse representation

?

Note: Make sure you select all of the correct options—there may be more than one!

CHECK

PREDICTION TARGET (1 point possible)

In the lab, we're trying to predict

- [ ] Revenue from click events
- [ ] Probability of a click event
- [ ] Click-through vs not click event

?

Note: Make sure you select all of the correct options—there may be more than one!

CHECK

OHE FEATURES (1 point possible)

In the lab, using the OHE method on the training data in Part (3c) creates a dictionary with:

- ( ) 23,328 features

- ○ 36,177 features
- ○ 233,286 features
- ○ 361,772 features

?

CHECK

FEATURE HASHING (1 point possible)

The feature hashing performed in the lab:

- ☐ Discards rare features
- ☐ Increases the number of features
- ☐ Requires calculating the OHE dictionary
- ☐ Causes feature collisions for certain observations

?

Note: Make sure you select all of the correct options—there may be more than one!

CHECK

SPARSE VECTORS (3 points possible)

In Part (1b) we use a sparse vector representation to efficiently store a one-hot-encoded (OHE) feature vector. Imagine that we have 1000 OHE features, and that for a particular data point, we have *s* non-zero OHE features.

If *s* = 10, how much smaller is the storage footprint of the sparse vector representation versus the dense representation (assume that all indices and values are stored as floats)?

- 100x
- 50x
- 10x
- they are the same size

?

If *s* = 500, how much smaller is the storage footprint of the sparse vector representation versus the dense representation (assume that all indices and values are stored as floats)?

- 100x
- 50x
- 10x
- they are the same size

?

Suppose we would like to compute a dot product between this feature vector and a dense vector, and assume *s* = 10. How many fewer scalar multiplications must we perform if we use a sparse vector representation versus a dense representation of the feature vector (assume we have random access to the entries of the dense vector)?

- 100x

- 50x
- 10x
- they are the same size

?

CHECK

HASHING (3 points possible)

In Part (5a) of the coding assignment we hashed the three sample points using numBuckets=4 and numBuckets=100. Complete the three statements below about these hashed features summarized in the following table using each answer once.

Name	Raw Features	4 Buckets	100 Buckets
sampleOne	[(0, 'mouse'), (1, 'black')]	{2: 1.0, 3: 1.0}	{14: 1.0, 31: 1.0}
sampleTwo	[(0, 'cat'), (1, 'tabby'), (2, 'mouse')]	{0: 2.0, 2: 1.0}	{40: 1.0, 16: 1.0, 62: 1.0}
sampleThree	[(0, 'bear'), (1, 'black'), (2, 'salmon')	{0: 1.0, 1: 1.0, 2: 1.0}	{72: 1.0, 5: 1.0, 14: 1.0}

With 100 buckets, sampleOne and sampleThree both contain index 14 due to _____.

?

It is likely that sampleTwo has two indices with 4 buckets, but three indices with 100 buckets due to _____.

?

With 4 buckets, sampleTwo and sampleThree both contain index 0 due to _____.

?

CHECK

SURVEY: LAB4 COMPLETION TIME (1 point possible)

How long did Lab FOUR take you to complete (in hours - decimals are OK)?

?

Please click "Check" to save your answers.

CHECK


Download on the
App Store


ANDROID APP ON
Google play