Principal Component Analysis

Latest Submission Grade
100%
1.
Question 1

Consider the following 2D dataset:

Which of the following figures correspond to possible values that PCA may return for u(1)u^{(1)}u(1) (the first eigenvector / first principal component)? Check all that apply (you may have to check more than one figure).

1 / 1 point
☑

Correct

The maximal variance is along the y = x line, so this option is correct.

☑

Correct

The maximal variance is along the y = x line, so the negative vector along that line is correct for the first principal component.

☐

☐

2.
Question 2

Which of the following is a reasonable way to select the number of principal components kkk?

(Recall that nnn is the dimensionality of the input data and mmm is the number of input examples.)

1 / 1 point
○

Choose kkk to be 99% of mmm (i.e., k=0.99∗mk = 0.99*mk=0.99∗m, rounded to the nearest integer).

Use the elbow method.

Choose kkk to be the largest value so that at least 99% of the variance is retained

Choose kkk to be the smallest value so that at least 99% of the variance is retained.

Correct

This is correct, as it maintains the structure of the data while maximally reducing its dimension.

3.
Question 3

Suppose someone tells you that they ran PCA in such a way that "95% of the variance was retained." What is an equivalent statement to this?

1 / 1 point

1m∑i=1m||x(i)−xapprox(i)||21m∑i=1m||x(i)||2≤0.05\frac{ \frac{1}{m} \sum_{i=1}^m ||x^{(i)}- x^{(i)}_{\rm approx}||^2}{\frac{1}{m} \sum_{i=1}^m ||x^{(i)}||^2} \leq 0.05m1∑i=1m||x(i)||2m1∑i=1m||x(i)−xapprox(i)||2≤0.05

1m∑i=1m||x(i)−xapprox(i)||21m∑i=1m||x(i)||2≥0.05\frac{ \frac{1}{m} \sum_{i=1}^m ||x^{(i)}- x^{(i)}_{\rm approx}||^2}{\frac{1}{m} \sum_{i=1}^m ||x^{(i)}||^2} \geq 0.05m1∑i=1m||x(i)||2m1∑i=1m||x(i)−xapprox(i)||2≥0.05

1m∑i=1m||x(i)−xapprox(i)||21m∑i=1m||x(i)||2≥0.95\frac{ \frac{1}{m} \sum_{i=1}^m ||x^{(i)}- x^{(i)}_{\rm approx}||^2}{\frac{1}{m} \sum_{i=1}^m ||x^{(i)}||^2} \geq 0.95m1∑i=1m||x(i)||2m1∑i=1m||x(i)−xapprox(i)||2≥0.95

○

1m∑i=1m||x(i)||21m∑i=1m||x(i)−xapprox(i)||2≥0.95\frac{\frac{1}{m} \sum_{i=1}^m ||x^{(i)}||^2}{\frac{1}{m} \sum_{i=1}^m ||x^{(i)}- x^{(i)}_{\rm approx}||^2} \geq 0.95m1 ∑i=1m||x(i)−xapprox(i)||2m1∑i=1m||x(i)||2≥0.95

Correct

This is the correct formula.

4.
Question 4

Which of the following statements are true? Check all that apply.

1 / 1 point

☐

Given only z(i)z^{(i)}z(i) and UreduceU_{\rm reduce}Ureduce, there is no way to reconstruct any reasonable approximation to x(i)x^{(i)}x(i).

☐

PCA is susceptible to local optima; trying multiple random initializations may help.

☑

Even if all the input features are on very similar scales, we should still perform mean normalization (so that each feature has zero mean) before running PCA.

Correct

If you do not perform mean normalization, PCA will rotate the data in a possibly undesired way.

☑

Given input data x∈Rnx \in \mathbb{R}^nx∈Rn, it makes sense to run PCA only with values of kkk that satisfy k≤nk \le nk≤n. (In particular, running it with k=nk = nk=n is possible but not helpful, and k>nk > nk>n does not make sense.)

Correct

The reasoning given is correct: with k=nk = nk=n, there is no compression, so PCA has no use.

5.

Question 5

Which of the following are recommended applications of PCA? Select all that apply.

1 / 1 point

☑

Data compression: Reduce the dimension of your input data x(i)x^{(i)}x(i), which will be used in a supervised learning algorithm (i.e., use PCA so that your supervised learning algorithm runs faster).

Correct

If your learning algorithm is too slow because the input dimension is too high, then using PCA to speed it up is a reasonable choice.

☑

Data compression: Reduce the dimension of your data, so that it takes up less memory / disk space.

Correct

If memory or disk space is limited, PCA allows you to save space in exchange for losing a little of the data's information. This can be a reasonable tradeoff.

☐

As a replacement for (or alternative to) linear regression: For most learning applications, PCA and linear regression give substantially similar results.

☐

Data visualization: To take 2D data, and find a different way of plotting it in 2D (using k=2).