S&DS 265 / 565

Introductory Machine Learning

PCA and Review

Tuesday, October 14

Yale

Plan for today

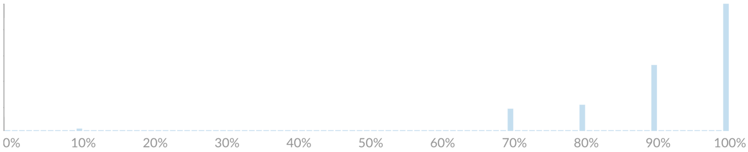
- Reminders
- Quick recap of PCA
- No new material
- Demo notebook
- Brief review for midterm

Quiz 2

Quiz Summary

Section Filter ▾ | Student Analysis | Item Analysis

Average Score	High Score	Low Score	Standard Deviation	Average Time
92%	100%	10%	1.16	15:11

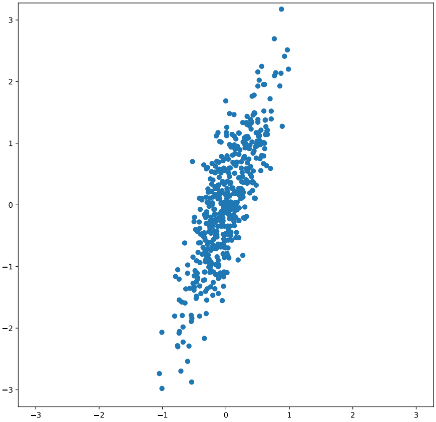


Reminders

- Assn 3 due today at midnight; Assn 4 out
- Midterm on Tuesday, October 19, in class
- “Closed book, notes, computer...”
- Allowed one $8\frac{1}{2} \times 11$ sheet of notes
- Practice midterm posted on Canvas (with solutions)
- Will go over practice midterm in review sessions
- Questions?

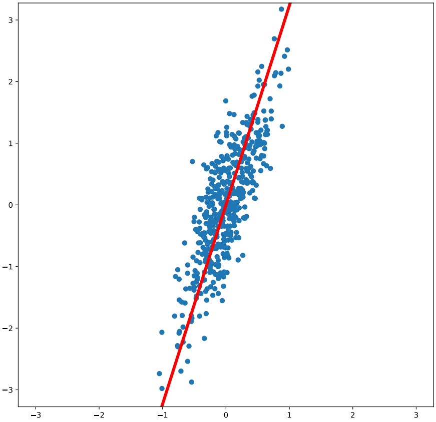
Principal Component Analysis (PCA)

PCA finds the directions of greatest variability in the data.

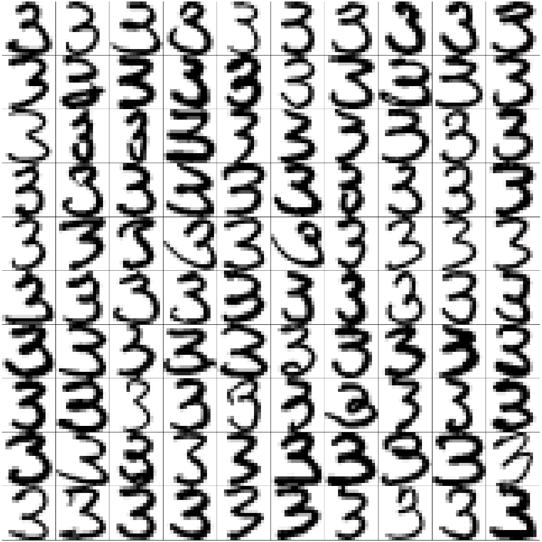


Principal Component Analysis (PCA)

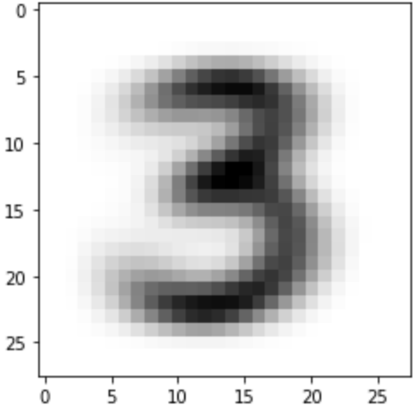
PCA finds the directions of greatest variability in the data.



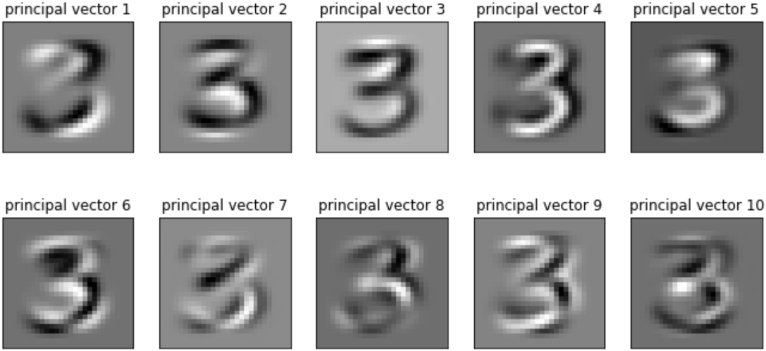
Handwritten Digits (3s)



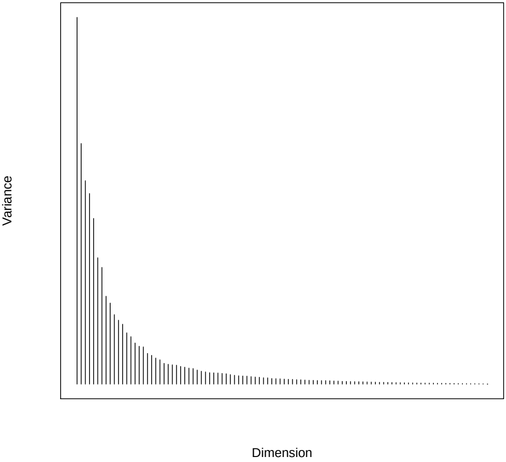
Handwritten Digits (3s) – Average



Handwritten Digits (3s) – Principal vectors

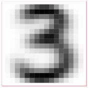
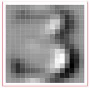
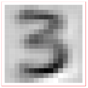


Handwritten Digits (3s) – PCA variance



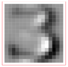
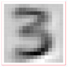
Handwritten Digits (3s)

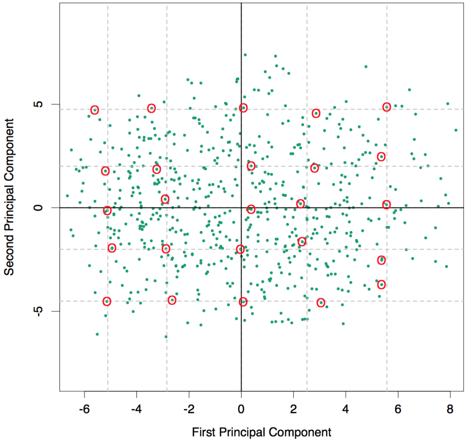
$$\hat{f}(\lambda) = \bar{x} + \lambda_1 v_1 + \lambda_2 v_2$$

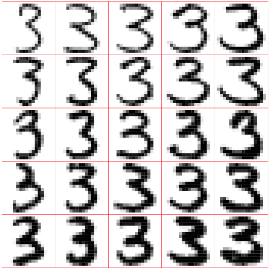
$=$  $+ \lambda_1 \cdot$  $+ \lambda_2 \cdot$  .

Handwritten Digits (3s) – Top 2 components

$$\hat{f}(\lambda) = \bar{x} + \lambda_1 v_1 + \lambda_2 v_2$$

$=$  $+ \lambda_1 \cdot$  $+ \lambda_2 \cdot$  .





PCA: Algorithm

1. Center the data: $x_i \mapsto x_i - \overline{x}$
2. Compute the $d \times d$ sample covariance $S = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T$
3. Find the first k eigenvectors of S
4. Project the data onto those k vectors

PCA: Algorithm

1. Center the data: $x_i \mapsto x_i - \frac{1}{n}\sum_{j=1}^{n} x_j = x_i - \overline{x}$

2. Compute the $d \times d$ sample covariance $S = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$. Note that

$$\frac{1}{n}\sum_i \left(x_{ij} - \overline{x}\right)^2$$

   is the sample variance of jth coordinate of data.

3. Find the first k eigenvectors of S,

$$\phi_1, \ldots, \phi_k \in \mathbb{R}^d, \qquad S\phi_j = \lambda_j \phi_j$$

4. Project the data onto those k vectors:

$$x_i \mapsto \overline{x} + (\phi_1^T x_i)\phi_1 + \ldots + (\phi_k^T x_i)\phi_k$$

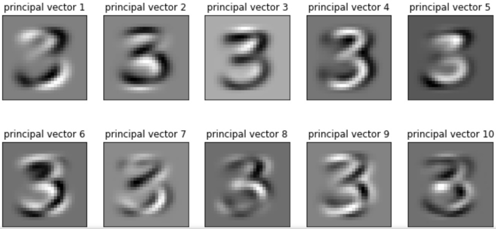
PCA: Algorithm

1. We can compute everything directly
2. Except for the eigenvectors
3. Let’s illustrate this in the demo notebook

Let’s go to the notebook

```
pca = PCA(num_components).fit(cimages)
principal_vectors = pca.components_
principal_vectors = principal_vectors.reshape((num_components, height, width))
pcs = pca.fit_transform(cimages)
capprox = pca.inverse_transform(pcs)
labels = ['principal vector %d' % (i+1) for i in np.arange(num_components)]
plot_images(principal_vectors, labels, height, width, int(num_components/5.), 5)
ratio = pca.explained_variance_ratio_.sum()
print('Variance explained by first %d principal vectors: %.2f%%' % (num_components, ratio*100))
```

```
Variance explained by first 25 principal vectors: 72.46%
```



Using PCA for classification or regression

- A combination of supervised learning and unsupervised learning
- Given data $\{x\}$ extract principal vectors and components
- Map each data point x_i to its principal components

$$z_i \equiv (x_i^T \phi_1, \ldots, x_i^T \phi_K)$$

- For labeled data $\{(x_i, y_i)\}$, now train a supervised learning algorithm using the transformed data $\{(z_i, y_i)\}$.

Example notebook

Flower Power: PCA and classification (30 points)

In this problem you will carry out principal components analysis and classification on the iris data. The task will be to reduce the dimension from four to two using PCA, and then to train logistic regression models on the projected data.

PCA: Summary

- PCA is an unsupervised method
- Finds directions of greatest variation in the data
- The directions are called the *principal vectors*; the weightings on the vectors are called the *principal components*
- The first few vectors may be interpretable
- Orthogonality makes interpretation difficult for the higher components
- Can be used for visualization or dimensionality reduction

Review

Week	Dates	Topics	Demos	Assignments & Exams	Slides and Readings
1	Sept 2	Course overview			Sept 2: Course overview
2	Sept 7, 9	Python and background concepts	CO Python elements CO Covid trends	Thu: Quiz 0	Data8 Chapters 3, 4, 5 Sept 7: Python elements Sept 9: Pandas and linear regression
3	Sept 14, 16	Linear regression and classification	CO Covid trends (revisited) CO Classification examples	Tue: CO Assn1 out	Sept 14: Regression concepts Notes on regression Sept 16: Classification Notes on classification
4	Sept 21, 23	Stochastic gradient descent	CO SGD examples	Tue: Quiz 1 Thu: Assn 1 in CO Assn2 out	Sept 21: Classification (continued) Sept 23: Stochastic gradient descent
5	Sept 28, 30	Bias and variance, cross-validation	CO Bias-variance tradeoff CO Covid trends (revisited) CO California housing		Sept 28: Bias and variance Sept 30: Cross-validation
6	Oct 5, 7	Tree-based methods	CO Trees and forests Visualizing trees	Tue: Assn 2 in CO Assn3 out	Oct 5: Trees Oct 7: Forests
7	Oct 12, 14	PCA and dimension reduction	CO PCA examples CO PCA revisited	Tue: Quiz 2 Thu: Assn 3 in; Assn 4 out	Oct 12: PCA
8	Oct 19	Midterm exam (in class)			Practice midterm (Canvas) Sample solutions (Canvas)