

S&DS 265 / 565
Introductory Machine Learning

PCA and Review

October 13

Plan for today

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

Quiz 3

Quiz Summary

Section Filter ▾

 Student Analysis

 Item Analysis

 Average Score

93%

 High Score

100%

 Low Score

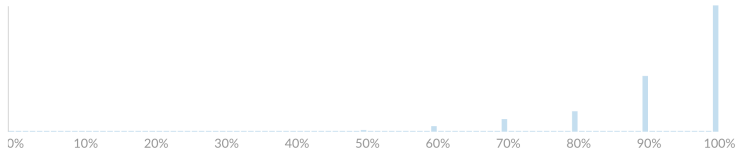
50%

 Standard Deviation

1.04

 Average Time

13:25

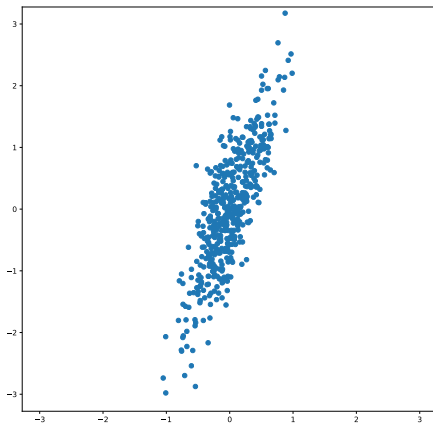


Reminders

- Assn 2 due today at midnight; Assn 3 out
- Midterm next Tuesday, October 18, in class
- “Closed book, notes, computer...”
- $8\frac{1}{2} \times 11$ sheet of notes, handwritten double-sided
- Practice midterms posted on Canvas (with solutions)
- Will go over practice exams in review sessions
- Questions?

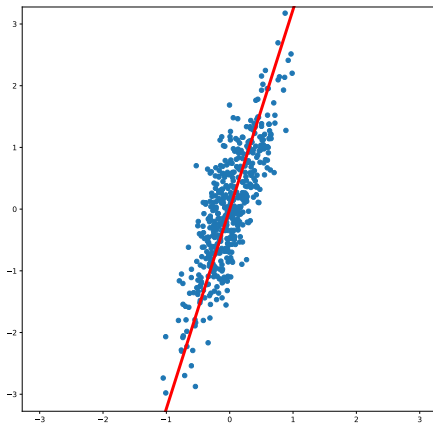
Principal Component Analysis (PCA)

PCA finds the directions of greatest variability in the data.

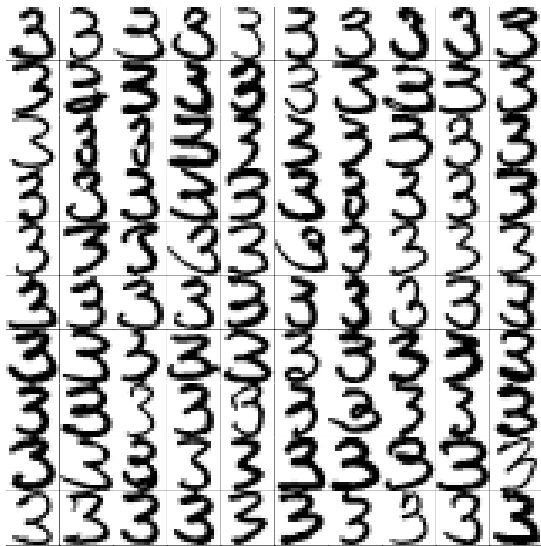


Principal Component Analysis (PCA)

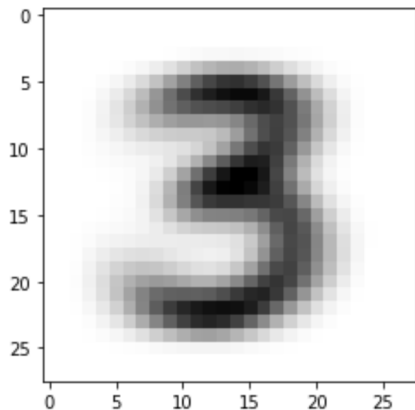
PCA finds the directions of greatest variability in the data.



Handwritten Digits (3s)

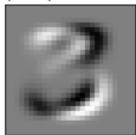


Handwritten Digits (3s) – Average

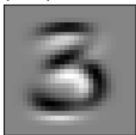


Handwritten Digits (3s) – Principal vectors

principal vector 1



principal vector 2



principal vector 3



principal vector 4



principal vector 5



principal vector 6



principal vector 7



principal vector 8



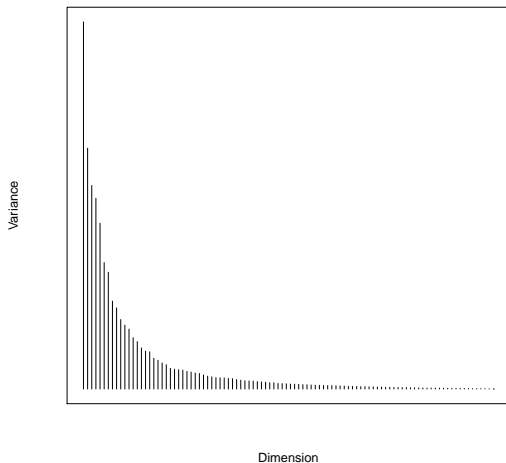
principal vector 9



principal vector 10



Handwritten Digits (3s) – PCA variance

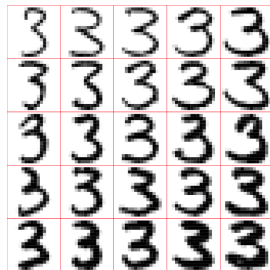
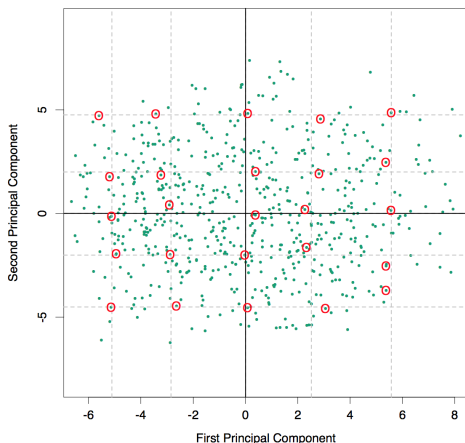


Handwritten Digits (3s)

$$\begin{aligned}\hat{f}(\lambda) &= \bar{x} + \lambda_1 v_1 + \lambda_2 v_2 \\ &= \boxed{\text{3}} + \lambda_1 \cdot \boxed{\text{3}} + \lambda_2 \cdot \boxed{\text{3}}.\end{aligned}$$

Handwritten Digits (3s) – Top 2 components

$$\begin{aligned}\hat{f}(\lambda) &= \bar{x} + \lambda_1 v_1 + \lambda_2 v_2 \\ &= \text{[Image of mean digit 3]} + \lambda_1 \cdot \text{[Image of } v_1 \text{]} + \lambda_2 \cdot \text{[Image of } v_2 \text{]}.\end{aligned}$$



PCA: Algorithm

- 1 Center the data: $x_i \mapsto x_i - \bar{x}$
- 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: $x_i \mapsto x_i - \frac{1}{n} \sum_{j=1}^n x_j = x_i - \bar{x}$
- 2 Compute the $d \times d$ sample covariance $S = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$. Note that

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

is the sample variance of j th coordinate of data.

- 3 Find the first k eigenvectors of S ,

$$v_1, \dots, v_k \in \mathbb{R}^d, \quad S v_j = \lambda_j v_j$$

- 4 Project the data onto those k vectors:

$$x_i \mapsto \bar{x} + (v_1^T x_i) v_1 + \dots + (v_k^T x_i) v_k$$

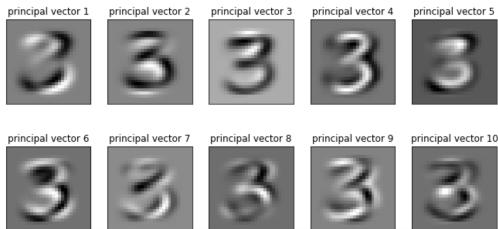
PCA: Algorithm

- 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

$$z_i \equiv (x_i^T v_1, \dots, x_i^T v_K)$$

- 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 & Tutorials	Lecture Slides	Readings and Notes	Assignments & Exams
1	Sept 1	Course overview		Sept 1: Course overview		
2	Sept 6, 8	Python and background concepts	Python elements Covid trends	Sept 6: Python elements Sept 8: Pandas and linear regression	Data8 Chapters 3, 4, 5	Thu: Quiz 1
3	Sept 13, 15	Linear regression and classification	Covid trends (revisited) Classification examples	Sept 13: Regression concepts Sept 15: Classification	ISL Sections 3.1, 3.2, 3.5 Notes on regression ISL Sections 4.3, 4.4 Notes on classification	Thu: Covid Assn1 out
4	Sept 20, 22	Stochastic gradient descent	SGD examples	Sept 20: Classification (continued) Sept 22: Stochastic gradient descent	ISL Section 6.2.2 ISL Section 10.7.2	Thu: Quiz 2
5	Sept 27, 29	Bias and variance, cross-validation	Bias-variance tradeoff Covid trends (revisited) California housing	Sept 27: Bias and variance Sept 29: Cross-validation	ISL Section 2.2 ISL Section 5.1	Thu: Assn 1 in Covid Assn2 out
6	Oct 4, 6	Tree-based methods	Trees and forests Visualizing trees Bagging operations	Oct 4: Trees Oct 6: Forests	ISL Sections 8.1, 8.2	Thu: Quiz 3
7	Oct 11, 13	PCA and dimension reduction	PCA examples PCA revisited Used for regression	Oct 11: PCA Oct 13: PCA and review	ISL Section 12.2	Thu: Assn 2 in Covid Assn3 out

Questions?