

# Correlated features

Our goal is to find ways to reduce the dimensionality of feature vectors.

Let's explore correlated features in the notebook on [Correlated features \(Unsupervised Correlated Features.ipynb\)](#).

# Principal Components: An alternate basis for our examples

Given that the features may be correlated

- We saw how changing the basis
- Can express the same examples
- In an alternate basis that is perhaps smaller

Let's formalize the notion of [alternate basis \(Unsupervised.ipynb#Alternate-basis\)](#).

# Principal components: introduction

We have seen how we can express the examples in  $\mathbf{X}$  in two coordinate systems

- The one with "original" features
- An alternate basis with "synthetic features"

Principal Components Analysis is the mechanism that we use

- To discover the new, alternate basis
- To find the feature values of examples, as measured in the alternate basis

Let's visit the [notebook section introducing PCA \(Unsupervised.ipynb#What-is-PCA\)](#).

# PCA: The math

The goal of PCA is to find a way of expressing examples  $\mathbf{X}$

- In a new basis  $V^T$
- With feature values  $\tilde{\mathbf{X}}$

$$\mathbf{X} = \tilde{\mathbf{X}}V^T$$

That is, we decompose  $\mathbf{X}$  into a product

- factorization of  $\mathbf{X}$

Let's go to the [notebook section on Matrix factorization \(Unsupervised.ipynb#PCA-via-Matrix-factorization\)](#) to explore how to factor  $\mathbf{X}$ .

# PCA: reducing the number of dimensions

Thus far

- Both the original basis and the new basis  $V$  have consisted of  $n$  basis vectors
- No information has been lost by the basis transformation

$$\mathbf{X} = \tilde{\mathbf{X}}V^T$$

If we are willing to lose some information

$$\mathbf{X}' \approx \mathbf{X}$$

we can achieve dimensionality reduction

- By an alternate basis  $(V^T)'$  of  $r \leq n$  basis vectors
- With synthetic feature vectors  $\mathbf{X}'$  of length  $r$



That is:  $\tilde{\mathbf{X}}'$  is a *reduced dimension* representation.

Questions to consider

- Which synthetic features to drop
- How many synthetic features to drop/keep

Let's go to the notebook section on [dimensionality reduction](#)  
([Unsupervised.ipynb#Dimensionality-reduction](#)).

# Transforming between original and synthetic features

We have thus far been concerned with the transformation

- From original features  $\mathbf{X}$
- To synthetic features  $\tilde{\mathbf{X}}$

We can also go in the opposite direction: from  $\tilde{\mathbf{X}}$  back to original features  $\mathbf{X}$

Let's go to the [notebook section on inverse transformation \(Unsupervised.ipynb#The-inverse-transformation\)](#).

# PCA in action

An example will hopefully tie together all the concepts.

Let's visit the [notebook section on PCA of small digits \(Unsupervised.ipynb#Example:-Reconstructing- \$x\$ -from- \$\tilde{x}\$ -and-the-principal-components\)](#).

# Choosing the number of reduced dimensions

Let's visit the [notebook section on PCA of MNIST \(Unsupervised.ipynb#MNIST-example\)](#) in order to see how the quality of approximation varies with the number of features in  $\tilde{\mathbf{X}}$

In [4]: `print("Done")`

Done