# **Correlated features**

Consider the following set of examples with 2 features



As you can see

-  $\mathbf{x}_2$  is perfectly correlated with  $\mathbf{x}_1$   $\mathbf{x}_2^{(\mathbf{i})} = 2*\mathbf{x}_1^{(\mathbf{i})}$ 

$$\mathbf{x}_2^{(\mathbf{i})} = 2 * \mathbf{x}_1^{(\mathbf{i})}$$

## Linear algebra

A way to conceptualize  $\mathbf{x}^{(i)}$ 

• As a point in the space spanned by unit basis vectors parallel to the horizontal and vertical axes.

$${f u}_{(1)}=(1,0)$$

$$\mathbf{u}_{(2)}=(0,1)$$

• With  $\mathbf{x^{(i)}}$  having exposure

$$\mathbf{x}_1^{(\mathbf{i})}$$
 to  $\mathbf{u}_{(1)}$ 

$$\mathbf{x}_2^{(\mathbf{i})}$$
 to  $\mathbf{u}_{(2)}$ 

So example  $\mathbf{x^{(i)}}$  is

$$\mathbf{x^{(i)}} = \sum_{j'=1}^2 \mathbf{x}_{j'}^{(i)} * \mathbf{u}_{(j')}$$

#### That is:

• Our feature space is defined by the basis vectors ("axes")

$$\mathbf{u}_{(1)} = (1,0)$$

$$\mathbf{u}_{(2)}=(0,1)$$

- $\mathbf{x^{(i)}}$  describes a point in the span of the basis vectors
  - $f x_1^{(i)}$  is the displacement of observation  $f x^{(i)}$  along basis vector  $f u_{(1)}$
  - $f x_2^{(i)}$  is the displacement of observation  $f x^{(i)}$  along basis vector  $f u_{(2)}$
- In general, for any length n vector of features

$$\mathbf{x^{(i)}} = \sum_{j'=1}^n ilde{\mathbf{x}}_{j'}^{(i)} * \mathbf{u}_{(j')}$$

One could easily imagine a different set of basis vectors to describe the feature space

- ullet For example: a rotation of basis vectors  ${f u}_{(1)},\ldots,{f u}_{(n)}$
- Let this alternate set of basis vectors be denoted by  $ilde{\mathbf{v}}_{(1)},\ldots, ilde{\mathbf{v}}_{(n)}$
- The basis vectors are mutually orthogonal

$$\tilde{\mathbf{v}}_{(1)} \cdot \tilde{\mathbf{v}}_{(2)} = 0$$

ullet The displacements  $\mathbf{x}_j^{(\mathbf{i})}$  need to be adjusted relative to the alternate basis

$$\mathbf{x^{(i)}} = \sum_{j'=1}^n ilde{\mathbf{x}}_{j'}^{(i)} * ilde{\mathbf{v}}_{(j')}$$

PCA is a technique for finding particularly interesting alternate basis vectors. The alternate basis is motivated by the fact that, for a given set of examples, there may be pairwise correlation among features. • If the correlation is *perfect* for some pair of features, they are redundant May drop one feature

Consider the set of examples above. Features 1 and 2 are perfectly correlated.

$$\mathbf{x}_2^{(\mathbf{i})} = 2*\mathbf{x}_1^{(\mathbf{i})}$$

We can create an alternate basis vector (no longer parallel to the axes)

$$\tilde{\mathbf{v}}_{(1)}=(1,2)$$

such that example  $\mathbf{x^{(i)}}$  is

$$\mathbf{x^{(i)}} = ilde{\mathbf{x}}_1^{(i)} * ilde{\mathbf{v}}_{(1)}$$

where  $ilde{\mathbf{x}}_1^{(\mathbf{i})} = \mathbf{x}_1^{(\mathbf{i})}$ 

That is,  $\mathbf{x}^{(i)}$  has exposure  $\tilde{\mathbf{x}}_1^{(i)}$  to the new, single basis vector.

So

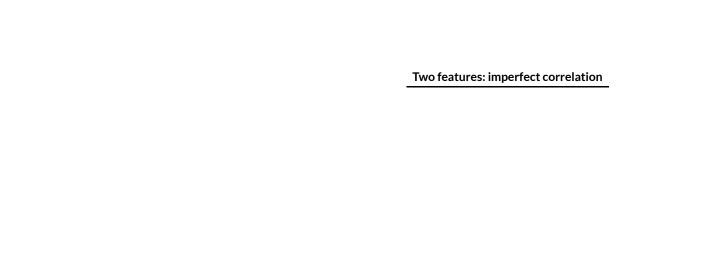
- Rather than representing  $\mathbf{x}^{(i)}$  as a vector with 2 features (in the original basis)
- We can represent it as  $\tilde{\mathbf{x}}^{(i)}$ , a vector with 1 feature (in the new basis)

This is the essence of dimensionality reduction

• Changing bases to one with fewer basis vectors

It is rarely the case for features to be perfectly correlated

Let's modify the set of examples just a bit.



We can still find an alternate basis of 2 vectors to perfectly describe the set of examples.

$$\mathbf{x^{(i)}} = \sum_{j'=1}^2 ilde{\mathbf{x}}_{j'}^{(i)} * ilde{\mathbf{v}}_{(j')}$$

ullet The dark black line is the first alternate basis vector  $oldsymbol{ ilde{v}}_{(1)}$ 

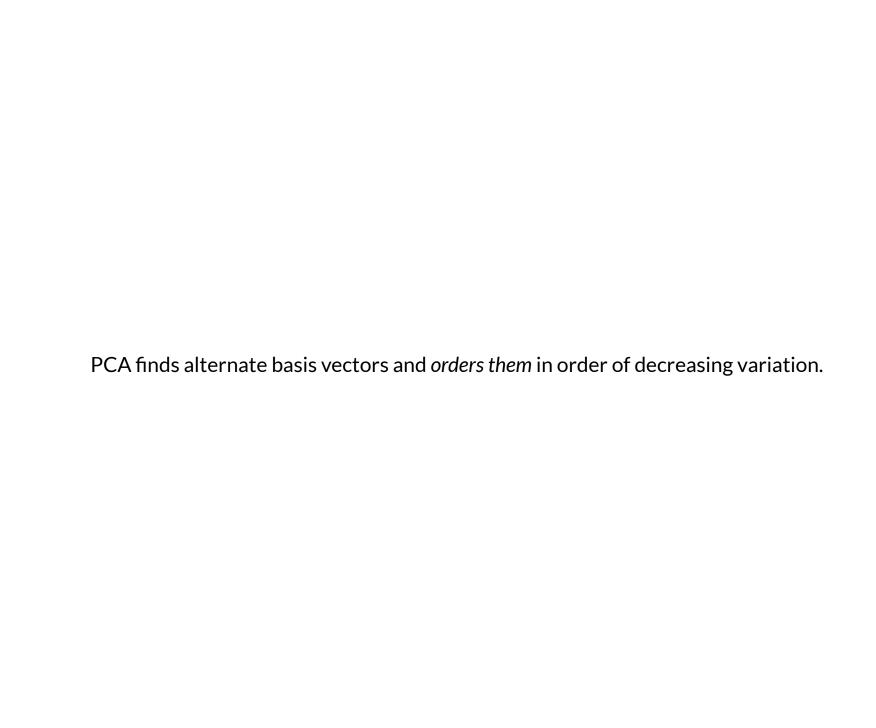


### As you can see:

- ullet The variation along  $ilde{\mathbf{v}}_{(1)}$  is much greater than that around  $ilde{\mathbf{v}}_{(2)}$
- $\bullet~$  Capturing the notion that the "main" relationship is along  $\tilde{\mathbf{u}}_{(1)}$

In fact, if we dropped  $ilde{\mathbf{v}}_{(2)}$  such that  $|| ilde{\mathbf{x}}||=1$ 

- The examples would be projected onto the line  $ilde{\mathbf{v}}_{(1)}$
- With little information being lost



## Subsets of correlated features

It may not be the case that a group of features is correlated across *all* examples

Consider the MNIST digits

- The subset of examples corresponding to the digit "1"
- Have a particular set of correlated features (forming a vertical column of pixels near the middle of the image)
- Which may not be correlated with the same features in examples corresponding to other digits

Thus, a synthetic feature encodes a "concept" that occurs in many but not all examples

We will present a method to discover "concepts"

- It may not necessarily be the pattern of features that corresponds to an entire digit
- It may be a partial pattern common to several digits
  - Vertical band (0, 1, 4, 7)
  - Horizontal band at top (5, 7, 9)

```
In [5]: print("Done")
```

Done