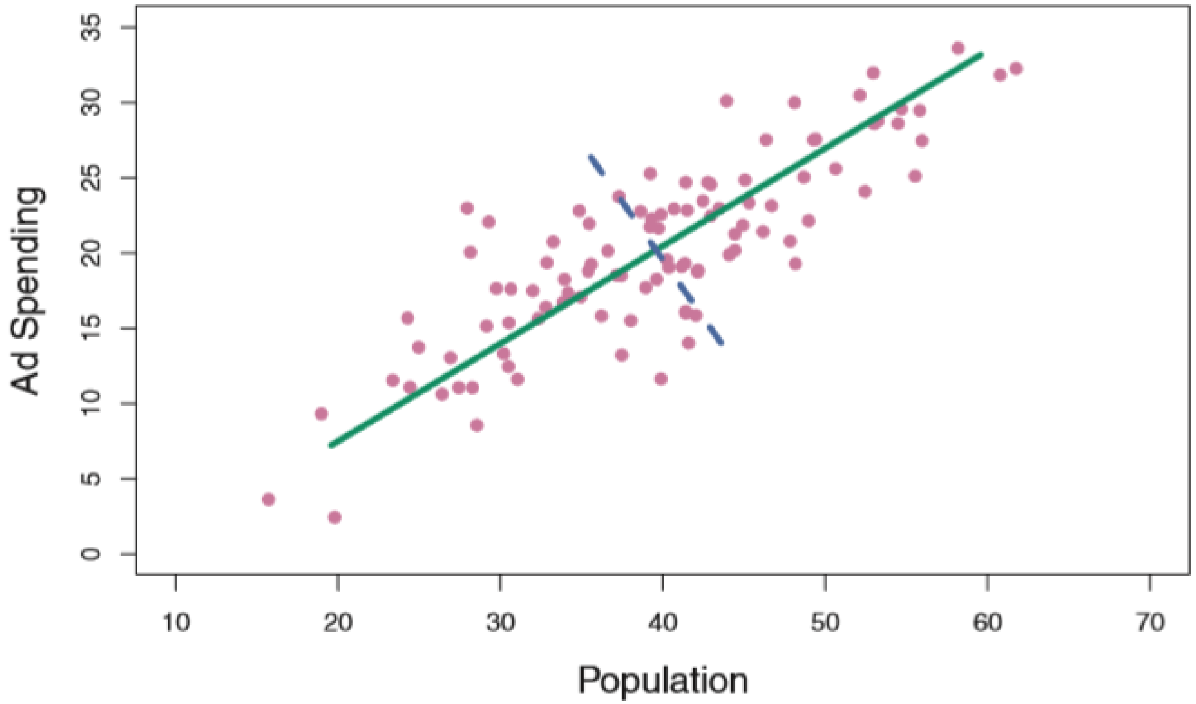
Statistical Machine Learning

Day 31

PCA, PCR, Partial Least Squares

PCA – find a lower dimensional representation of data



$$Z_1 = 0.839 \times (\text{pop} - \overline{\text{pop}}) + 0.544 \times (\text{ad} - \overline{\text{ad}}).$$

$$Z_2 = 0.544 \times (\text{pop} - \overline{\text{pop}}) - 0.839 \times (\text{ad} - \overline{\text{ad}}).$$

FIGURE 6.14. *The population size (pop) and ad spending (ad) for 100 different cities are shown as purple circles. The green solid line indicates the first principal component, and the blue dashed line indicates the second principal component.*

PCA – Project data onto lower dimensional space

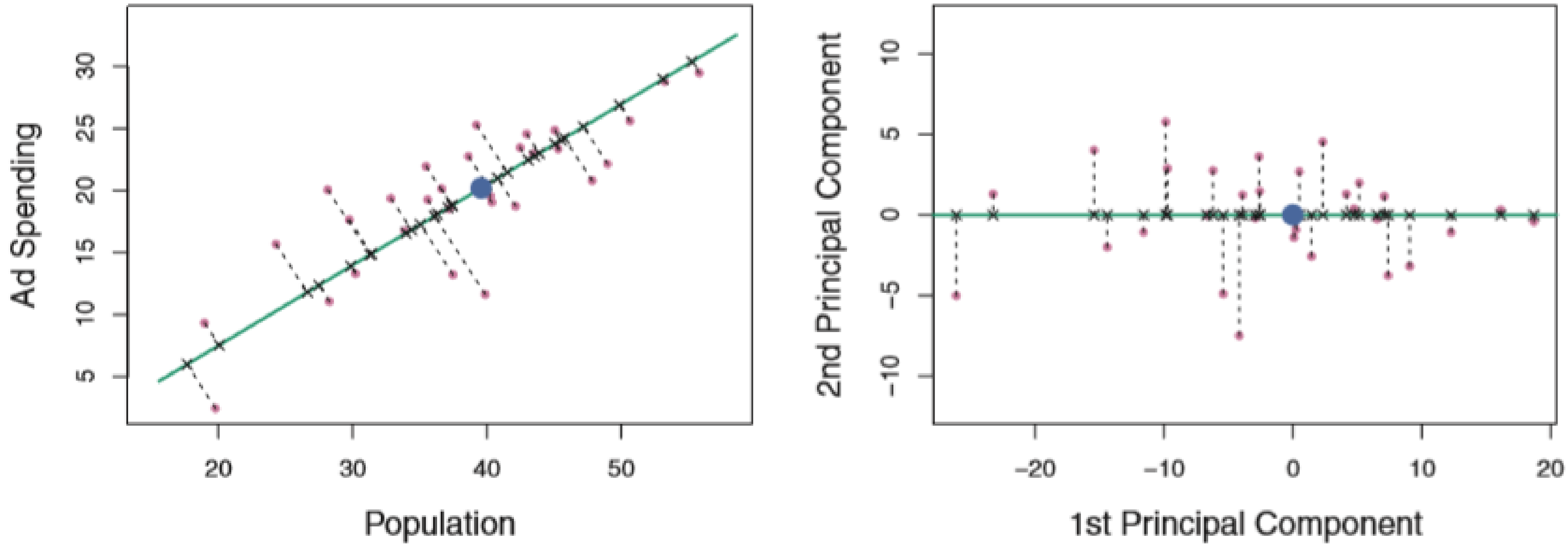


FIGURE 6.15. *A subset of the advertising data. The mean* pop *and* ad *budgets are indicated with a blue circle.* Left: *The first principal component direction is shown in green. It is the dimension along which the data vary the most, and it also defines the line that is closest to all n of the observations. The distances from each observation to the principal component are represented using the black dashed line segments. The blue dot represents* $(\overline{\text{pop}}, \overline{\text{ad}})$. Right: *The left-hand panel has been rotated so that the first principal component direction coincides with the x-axis.*

PCA – find a lower dimensional representation of data

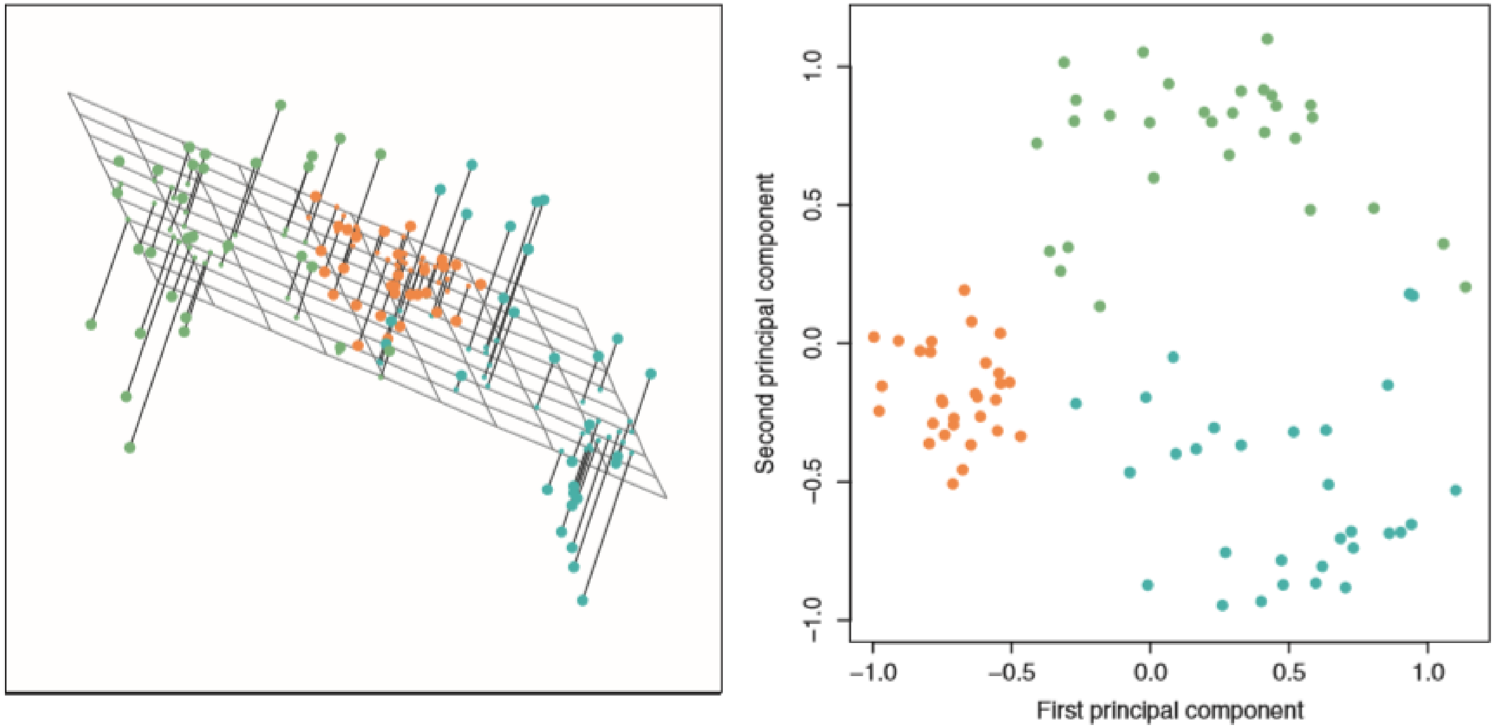
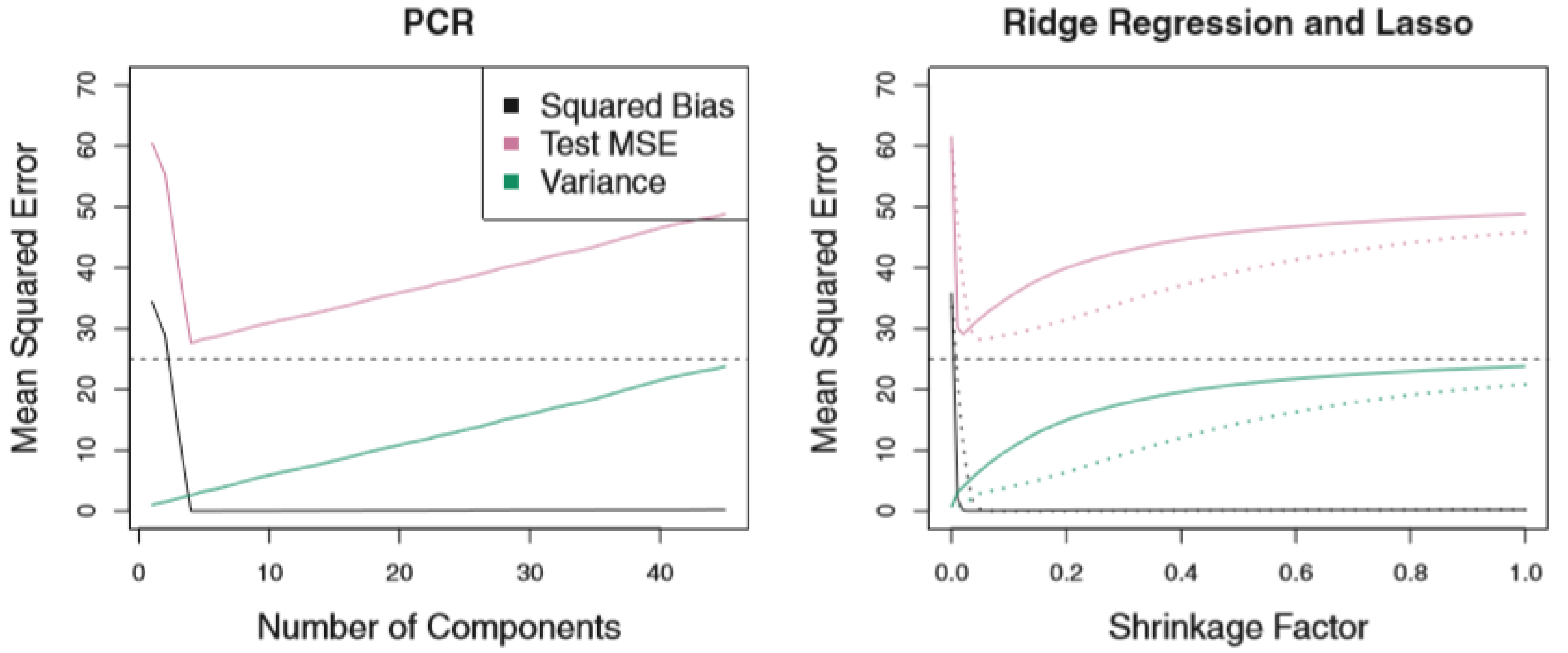
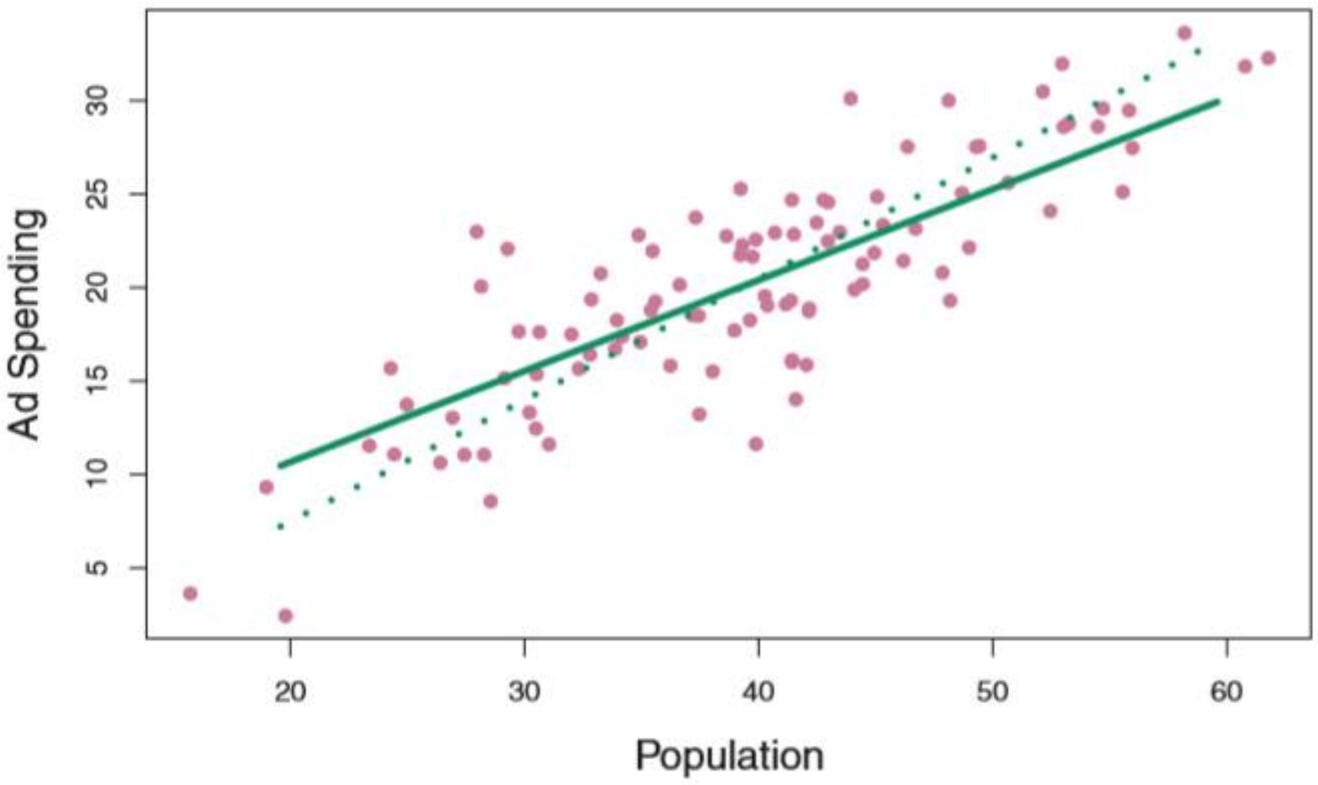


FIGURE 10.2. *Ninety observations simulated in three dimensions.* Left: *the first two principal component directions span the plane that best fits the data. It minimizes the sum of squared distances from each point to the plane.* Right: *the first two principal component score vectors give the coordinates of the projection of the 90 observations onto the plane. The variance in the plane is maximized.*

PCR – linear regression on the first k principal components



Partial Least Squares – improved PCR to find components most related to Y

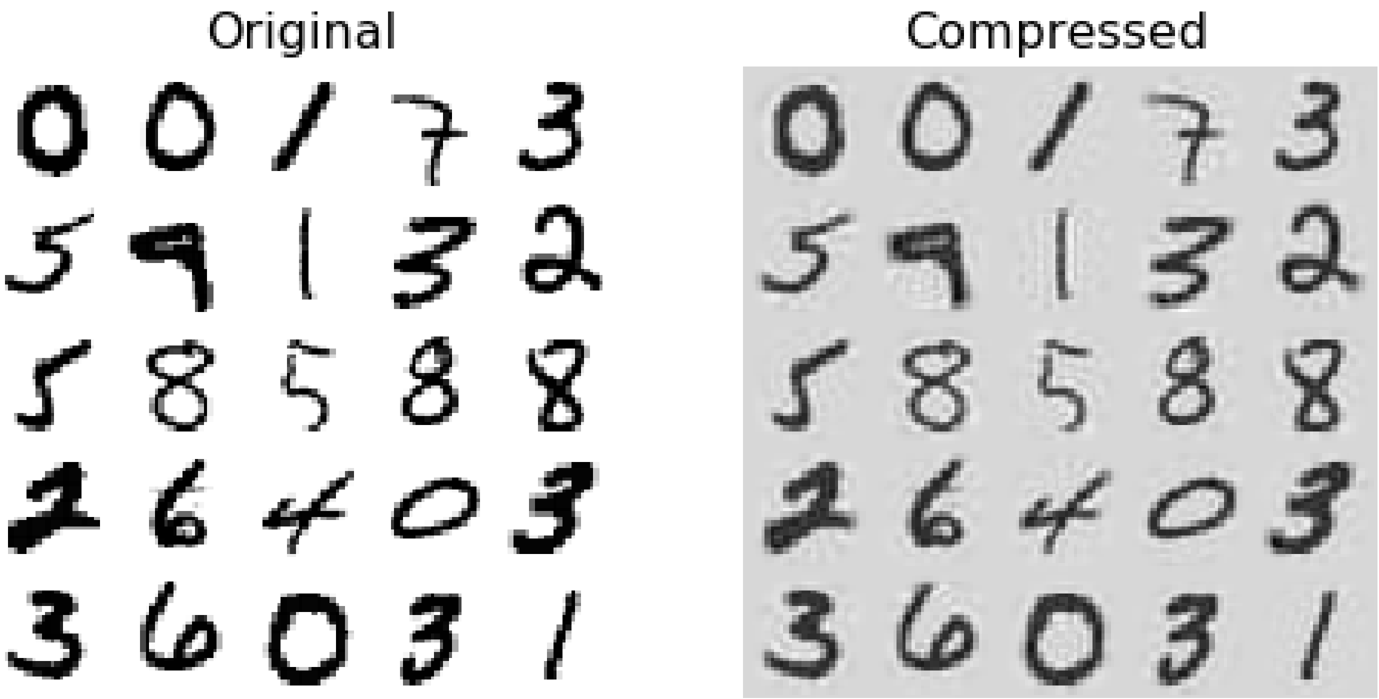


Summary

- PCA – unsupervised technique for finding a lower dimensional representation of the data
- PCR – regression (linear, logistic, generalized) on the first k principal components from PCA
- PLS – like PCR but the PCA part is supervised

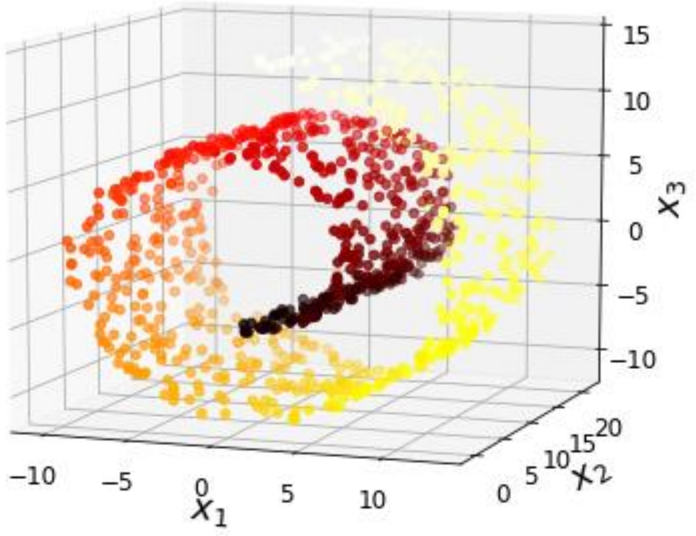
Often PCR and PLS have similar performance to ridge regression. PCA by itself is a very powerful technique…

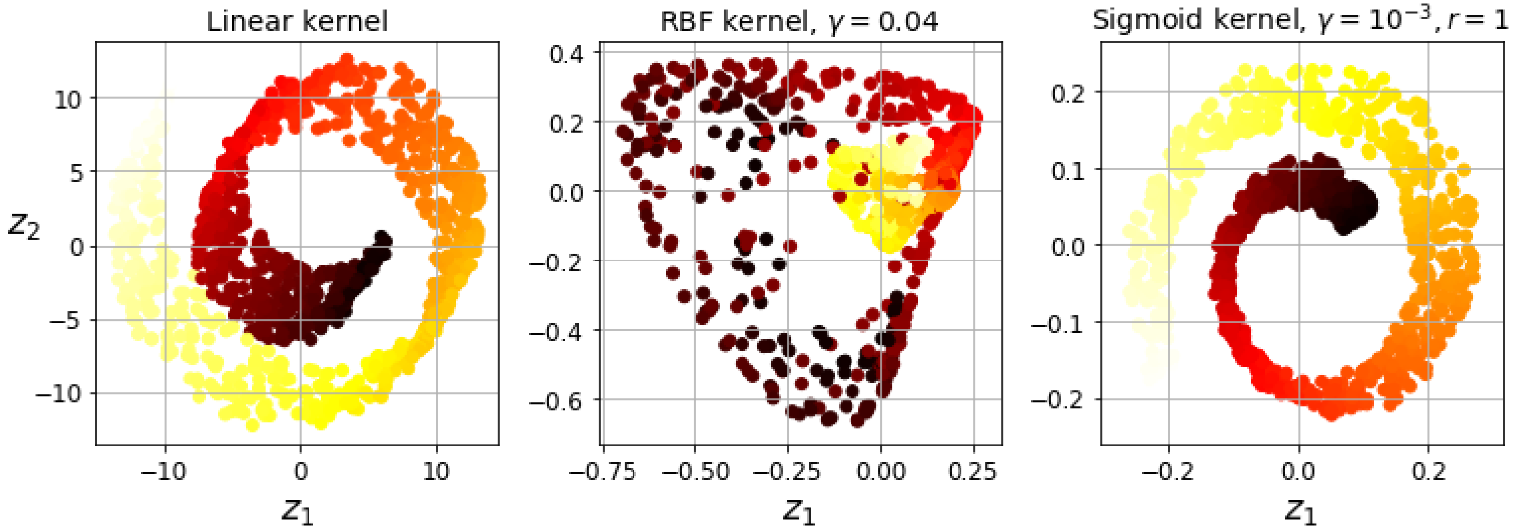
Lower dimensional representation



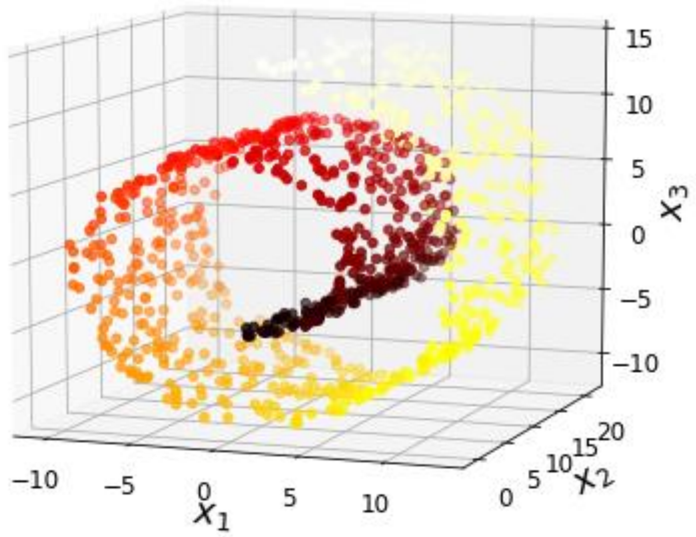
The next few images are from "Hands on Machine Learning" available for free to you via the library's safari books

Unrolling the “swiss roll” via lower dimensional representation that is not necessarily linear.

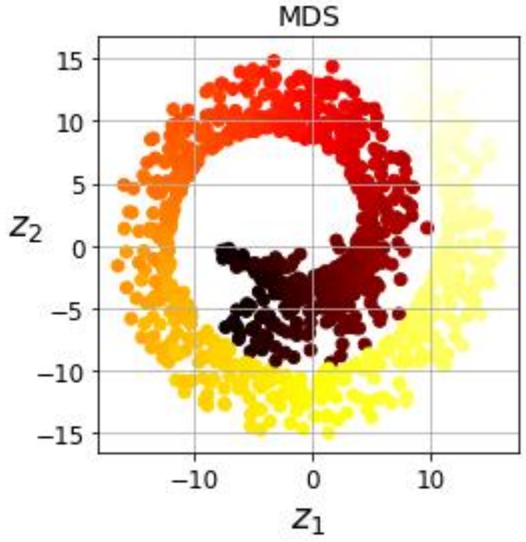




Multi-Dimensional Scaling (MDS)



Finds the k-dimensional space that best preserves the “distance” between data points.



Semi-Supervised Learning

Suppose we have only a small number of datapoints with labels (i.e. observations of **X** and Y) and a large number of unlabeled observations of **X.**

Cluster all observations of **X,** regardless of whether Y is available. (unsupervised learning)

Classify by the closest labeled example of **X**

or **estimate** by e.g. the mean of nearest labeled examples **of X**

(supervised learning)