8.2.3 Methods that use Linear Combinations of Attributes

A third approach to deal with multiple attributes, some of them correlated, is to create new linear combinations of predictive attributes, as explained in Chapter 3 for principal components analysis. These can be used for the linear regression instead of the original attributes.

8.2.3.1 Principal Components Regression

Principal components regression (PCR) creates linear combinations of predic- tive attributes. The first principal component – the first linear combination – is the one that captures the most variance of all the possible linear combinations. The following principal components are those ones that capture the most remaining variability while being uncorrelated with all previous principal components. PCR defines the principal components without evaluating how correlated the principal components generated are with the target attribute. The principal components are used as predictive attributes in the formulation of the multivariate linear regression problem.

8.2.3.2 Partial Least Squares Regression

Partial east squares (PLS) regression starts by evaluating the correlation of each predictive attribute with the target attribute. The first principal com- ponent is a linear combination of the predictive attributes, with the weights for each defined according to the strength of their univariate effect on the target attribute. The process follows the one described for PCR. PLS gives similar results to PCR but using fewer principal components. This is due to the process of measuring the correlation of the principal components with the target attribute.

8.3 Technique and Model Selection

- Random train/test split: This is a resampling method. ...
- Cross validation: It is a very popular resampling method for model selection. ...
- Bootstrap: This is also a resampling method, and can be performed like random train/test split or cross validation.

So whenever we have a new predictive task, we face the problem of which pre- dictive technique to choose. We can reduce the alternatives by selecting among those techniques known for their good performance in predictive tasks. But even so, we will still have a large number of options to choose from. The choice of the technique is influenced by our requirements, which can depend on:

- **Memory**
  - *Memory needed for the technique:* For data stream applications, where new data continuously arrive and classification models need to be automati- cally updated, if the technique is implemented in a portable device, it must fit the available device memory.
  - *Memory needed for the induced model:* The memory necessary to store the model is particularly important when the model is implemented on a small portable device and needs to be stored in a small memory space.

- **Processing cost**
  - *Technique processing cost:* This measure is related with the computational cost of applying the technique to a data set in order to induce a classification model. This is particularly relevant in data stream applications, where the model induction or update must be fast.
  - *Model processing cost:*This is the time the model takes, given the values of the predictive attributes of an object, to predict a class label of an object, which is relevant for applications that need a fast output, such as autonomous driving.
- **Predictive performance**
  - *Technique predictive performance:* This measure estimates the predictive performance of the models induced by the technique. This is the main performance measure and often the main aspect taken into account for technique selection. It is usually the average predictive performance of several models induced by the technique for different samples of a data set. It depends on the predictive performance of the models induced by the technique.
  - *Model predictive performance:* This estimates the predictive performance of a classification model induced by a technique for the classification of new data.
- **Interpretability**
  - *Technique interpretability:* How easy it is to understand the knowledge represented by the models induced by the technique?
  - *Model interpretability:* How easy it is for a human to understand the knowledge represented by a particular model?