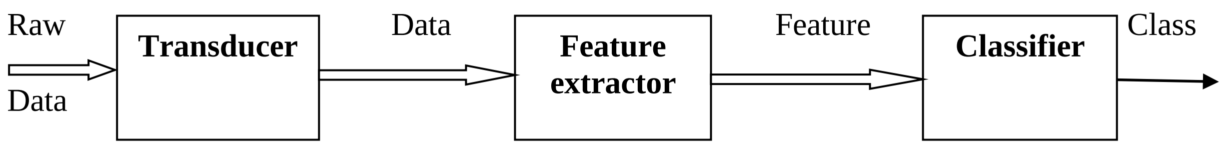
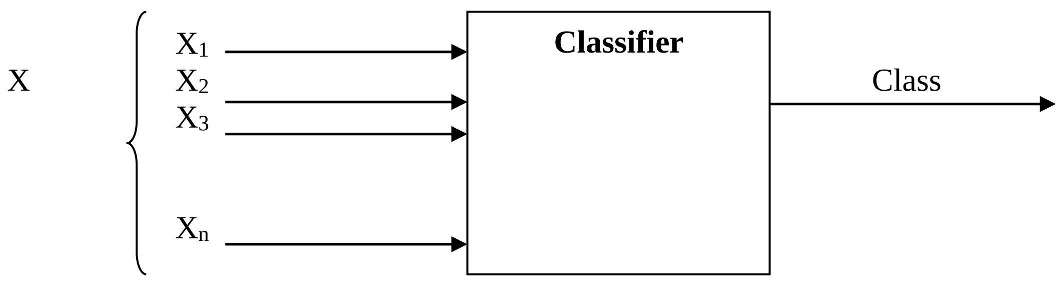
Classification Model for Single-Layer Perceptron Classifiers

One of the most useful tasks that can be performed by networks of interconnected nonlinear elements introduced is pattern classification. A ***pattern*** is the quantitative description of an object or event. The classification may involve spatial patterns as pictures, video images of ships, weather maps, fingerprints, characters and temporal patterns as speech signals, signals vs time as produced by sensors, ECG and seismograms. Temporal patterns usually involve ordered sequences of data appearing in time. The goal of pattern classification is to assign a physical object or event to one of the pre-specified class/category.

While some of the classification tasks can be learned easily, the growing complexity of the human environment and technological progress has created classification problems that are diversified and also difficult. As a result, the use of various classifying aids became helpful. Reading and processing bank checks exemplifies a classification problem that can be automated. It obviously can be performed by a human worker; however, machine classification can achieve much greater efficiency. Eventually, machine classification came to maturity to help people in their classification tasks. The electrocardiogram waveform, biomedical photograph, or disease diagnosis problem can nowadays be handled by machine classifiers. Other applications include fingerprint identification, patent searches, radar and signal detection, printed and written character classification, and speech recognition.

The classifying system consists of an input transducer providing the input pattern data to the feature extractor. Typically, inputs to the feature extractor are sets of data vectors that belong to a certain category. Assume that each such set member consists of real numbers corresponding to measurement results for a given physical situation. Usually, the converted data at the output of the transducer can be compressed while still maintaining the same level of machine performance. The compressed data are called **features**. The feature extractor at the input of the classifier performs the reduction of dimensionality. The feature space dimensionality is postulated to be much smaller than the dimensionality of the pattern space. The feature vectors retain the minimum number of data dimensions while maintaining the probability of correct classification, thus making handling data easier.





The classifier input components can be represented as a vector x. The classification at the system's output is obtained by the classifier implementing the decision function $i_0(x)$. The discrete values of the response i, are 1 or 2 or . . . or R. The responses represent the categories into which the patterns should be placed. The classification (decision) function is provided by the transformation, or mapping, of the n-component vector x into one of the category numbers i, as $i = i_0(x)$,

$$\text{where } x = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ : \\ X_n \end{pmatrix}$$

There are two simple ways to generate the pattern vector for cases of spatial and temporal objects to be classified. First, each component xi of the vector **x** = $[xl\ x2\ .\ .\ .\ xn]^t$ is assigned the value 1 if the i'th cell contains a portion of a spatial object; otherwise, the value 0 (or - 1) is assigned. In the case of a temporal object being a continuous function of time t, the pattern vector may be formed at discrete time instants t_i by letting xi = f (ti), for i = 1, 2, . . . , n.

Classification can often be conveniently described in geometric terms. Any pattern can be represented by a point in n-dimensional Euclidean space E^n called the pattern space. Points in that space corresponding to members of the pattern set are n-tuple vectors x. A pattern classifier maps sets of points in E^n space **into** one of the numbers i_0 = 1, 2, . . . , ***R,*** as described by the decision function. The sets containing patterns of classes 1, 2, . . . , ***R*** are denoted here by X_1, X_2 , . . . , X_R, respectively.

The regions denoted by X_i are called decision regions. Regions are separated from each other by decision surfaces. We shall assume that patterns located on decision surfaces do not belongs to any category. The decision surfaces in two-dimensional pattern space $\boldsymbol{E^2}$ are curved lines.

Discriminant functions: Let us assume that the classifier has already been designed so that it can correctly perform the classification tasks. During the classification step, the membership in a category needs to be determined by the classifier based on the comparison of R discrirn*inant functions* $g_l(x)$, $g_2(x)$, . . . , $g_R(x)$; computed for the input pattern under consideration. It is convenient to assume that the discriminant functions $g_i(x)$ are scalar values and that the pattern x belongs to the i'th category if and only if

$$g_i(x) > g_j(x), \text{ for } i, j = 1, 2, . . . , R, i \neq j$$

So, within the region X_i, the ith discriminant function will have the largest value. This maximum property of the discriminant function $g_i(x)$ for the pattern of class ***i*** is fundamental, and it will be subsequently used to choose, or assume, specific forms of the $g_i(x)$ functions.

The discriminant functions' $g_i(x)$ and $g_j(x)$ for contiguous decision regions Xi and Xj define the decision surface between patterns of classes i and j in $\boldsymbol{E^n}$ space. Since the decision surface itself contains patterns x without membership in any category, it is characterized by $g_i(x)$ equal to $g_j(x)$. Thus, the decision surface equation is $g_i(x) - g_j(x) =0$

Once a general functional form of the discriminant functions has been suitably chosen, discriminants can be computed using a ***priori*** information about the classification of patterns, if such information is available. In such an approach, the design of a classifier can be based entirely on the computation of decision boundaries as derived from patterns and their membership in the classes. Once a type of discriminant function has been assumed, the algorithm of learning should result in a solution for the initially unknown coefficients of discriminant functions, provided the training pattern sets are separable by the assumed type of decision function. For study of such adaptive, or trainable, classifiers, the following assumptions are made:

1. The training pattern set and classification of all its members are known, ie. the training is supervised.
2. The discriminant functions have a linear form and only their coefficients are adjusted in the training procedure.

We select an input data vectors whose correct class is known. These vectors will be referred to as class prototypes or exemplars. The classification problem will be one of finding decision surfaces, in n-dimensional space, that will enable correct classification of the prototypes and will afford some degree of confidence in correctly recognizing and classifying unknown patterns that have not been used for training.