BASIC CONCEPTS

Pattern recognition is a process by which some input is measured, analyzed, and then classified as belonging to a certain class.

Pattern: A pattern is assumed to have certain properties or attributes which distinguish it from other patterns.

Measurements: measurements are taken of a pattern, which (should) reflect either directly or indirectly the attributes associated with the pattern. One of the steps in pattern recognition is to assess to what extent a given measurement is relevant or helpful, as opposed to irrelevant or useless, to the classification task.

Features: features are functions of measurements,

The process of transforming measurements into features is intended to facilitate the classification process:

- By reducing the dimensionality of the problem: $n<m$. Fewer dimensions are easier to visualize, easier to understand, are normally associated with learning fewer parameters, and will typically lead to more robust classifiers.
- Generate features that can make the patterns more distinguishable: each measurement may only very weakly distinguish different pattern classes, and it may be that some function f() of the measurements can yield a feature which is much more discriminating

Note: Feature extraction never adds any extra information, just transforms the information in the measurements.

Example: Image Classification:

Classes: A class is a particular pattern, or possibly a group of patterns that are similar or equivalent in some sense. In a given problem, the set of classes C is defined as:

Whether we know the classes, or what we know about the classes, or whether we even know the number of classes K, will depend on the kind of pattern recognition problem to be solved. For example, for a clustering task, the classes might be unknown.

Assumption: members of samples of the class share some attributes, which will eventually help us classify patterns.

Describing classes: For a pattern recognition to classify patterns, it will need some representation of the classes to reason on. There are four main types of class descriptions:

- : an idealized representation or notion of the "essence" of the class. Advantage: pretty clear definition of the class, disadvantage: no variability.
- : a generalization of the prototype, in which the class has a known shape (rectangular or elliptical, say), where the shape is described in some number of parameters (ellipse centre, rotation, and axis lengths, for example). Advantage: more flexible than the prototype, disadvantage: still assumes a particular shape beforehand.
- : such that we have some description of the probability of a class member having a particular set of measurements or features. Advantage: Quite powerful, disadvantage: the distribution might be unknown and difficult to infer.
- : such that a set of given samples (many apples, or tigers, or bicycles) directly characterizes the class. When such samples are given the representation is highly convenient, since nothing further needs to be done to describe the class. Disadvantage: storage and computational challenges, since all of the data need to be saved and searched every time for classification of a single sample.

Variability in Classes: Patterns belonging to a particular class do not need to be (and most like will not be) the same. There are two sources that can cause this variability in classes:

- Every class will consist of members which differ in some way. For example, the "Fruit" class contains all manners of variability in colour, size, and shape; the "Apple" class is much more specific, but apples do come in different colours and patterning; the "Gala Apple" class is even more specific, but still will have apples of different sizes or some further differences in the texture.
- Noise or random variations in measurement: Every measurement involves some sort of the physical process which will be subject to error, such as thermal noise in electronics, or quantization noise in converting an analogue signal to a digital representation.

Therefore, A class typically spans a region within feature space to account for differences in measurement and feature values amongst its members.

Classification:
Eventually, what we really care for is can this information (measurements and features) be used to perform classification. For this, we need a classifier. A classifier is some function $g()$, possibly analytical (i.e., an equation) or a computer algorithm, which assigns a class label to a given feature (or pattern to be general)

There is a strong relationship between feature extraction and classification. For example: good features allow for simple classifiers (e.g., best feature is just the class label!) Complex classifiers compensate for features that are not linearly separable.

Classifier Learning: How do we define the function $g()$? We can use methods for learning g() on the basis of the class representation we choose (the four class representations described above).

Classifier Testing:
After training, we want to check how well the classifier performs. We can check training or test accuracy for that.
Training accuracy: As the name implies, it determines how well a classifier can assign correct class labels to samples it was already trained on.
Test accuracy: determine how effectively a classifier can assign correct classes to test samples it has not been exposed to.

Types of Classification Problems:

- *Model Based*: Probability model is known for each class: This is the most information we can hope to have regarding a pattern recognition problem, in that we are told the behaviour of the measurements for each of the pattern.
    - Typical in cases where the physical process is known and provides a probability model (e.g., Gaussian noise), or reasonable assumption can be made about the probabilistic behavior (e.g., car arrival time as a Poisson process)
    - Statistical decision theory may be used to find optimal classification in the sense of minimizing probability of error.
- *Supervised Learning*: Model is not known; labelled data are available.
    - Although we do not have an exact description of the problem, as in the previous case, we are given labelled data. For example, data pairs of the form:
    - Two possible approaches: i) learn empirical probability model based on samples, and ii) derive classifiers directly from distribution of samples in feature space.
- *Semi-Supervised Learning*: The model is unknown; some data are known and some unknown
    - Semi-supervised problems appear commonly in cases where we wish to use a small set of samples that have been manually labelled for classification.
    - Should not ignore unlabeled data when learning a classifier.
- *Unsupervised Learning*: model and class labels both are unknown.
    - Need to not only determine the definition of each class, but also determine the number of classes!
    - Commonly referred to as a clustering problem (or density estimation to be general).
    - Approach is to look for naturally occurring order, groupings, or clusters in the data (with the assumption that members of a class share some attributes).

Regression Vs Classification:

- *Regression*: A pattern recognition problem where the output is not a discrete number/label of a class, but rather a real number.
- *Classification*: In this pattern recognition problem, the output of the function/model is discrete, usually class/category labels.

Bias-Variance Trade-off: