- What is recognition?
- How do we recognize things?
- Bayesian categorization approach.

1 What is recognition

Recognition, (from RE - Cognize) is a term used to describe an ability to identify things based on prior knowledge.

1.1 Recognition is a hard task

The problem of recognition is hard due to many sources of variability:

1. **Each object in the class exibits variability**. What is a dog or a chair?

2. **There is variability in lighting between instances of the same object**. Simple intensity matching will not work.

3. **Variation in pose**. Orientation and position of object with respect to camera or observer.

1.2 What enables recognition

Most object have specific characteristic features. Features can be visual (color, shape, texture, motion) or non-visual (smell, taste, sound, etc.).

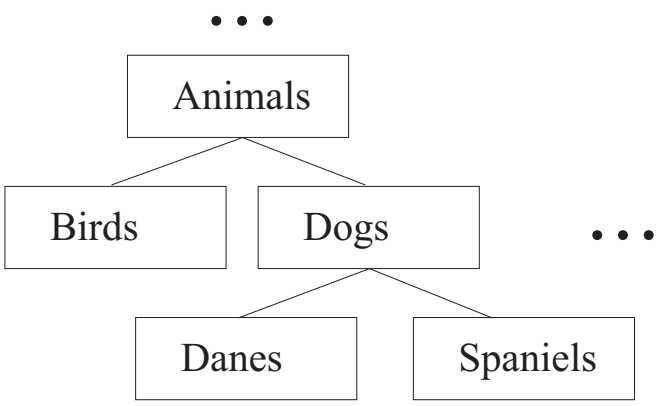

Figure 1: Object hierarchies

2 How do we recognize things

A question that can be asked about recognition, is whether it is based on function or categoryzation. There are two basic approaches to this, - **affordance** and **categorization**.

2.1 Affordance

Affordance, proposed by Gibson, postulates two assumptions:
• Memory is not needed for recognition. Just the "bottom-up", geometric information obtained from a scene is sufficient.
• Objects are recognized by their function, as suggested by their geometry. For example, a chair can be recognized as a flat surface of correct height that can support human weight, etc.

However, some objects are similar in geometric properties, but different in function (mailbox vs. trash can).

2.2 Categorization

Categorization is approach advocated by most of the scientific community. It is based on the following principles:
• Entities are learned from examples (by label association).
• The examples can be generalized to categories or classes.

Each object belongs to multiple categories that have hierarchical structure (Figure 1) .

"Basic Category" is the level of hierarchy that is recognized the fastest. This natural category depends on the observer and his/her expertise with the subject area.

3 Bayesian categorization approach

How to classify patterns or features? We adopt Baesian approach. Let there be two classes C_1 and C_2. Let x be a feature vector. For simplisity we assume that x is scalar. For example categories can be male and female and the feature variable x is the person's height.

We can define probability distribution of feature x, given the class as $P(x|C_1)$ and $P(x|C_2)$. These probabilities are infered from the data. Priors, $P(C_1)$ and $P(C_2)$ represent relative frequencies of the two classes. Using Bayes theorem:

$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x)}$ and $P(C_2|x) = \frac{P(x|C_2)P(C_2)}{P(x)}$

If there are only two possible classes, like in this example, we can rewrite $P(C_1|x)$ as:

$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1)+P(x|C_2)P(C_2)} = \frac{1}{1+\frac{P(x|C_2)P(C_2)}{P(x|C_1)P(C_1)}}$

Define $z = ln(\frac{P(x|C_1)P(C_1)}{P(x|C_2)P(C_2)})$, so that:

$P(C_1|x) = \frac{1}{1+e^{-z}}$

z is called log-likelihood ratio, and the conditional probability is a logistic function of z. It is plotted on figure 2.

Sometimes, the posterior distribution $P(C_1|x)$ can be learned directly from data, but first we consider a special case, when $P(x|C_1)$ is a Gaussian distribution. The class conditional probabilities can be arbitrary, but Gaussian distribution model allows simple mathematical framework.

Let $P(x|C_1) = \frac{1}{\sqrt{2*\pi}\sigma}e^{-\frac{1}{2}(\frac{x-\mu_1}{\sigma})^2}$ and $P(x|C_2) = \frac{1}{\sqrt{2*\pi}\sigma}e^{-\frac{1}{2}(\frac{x-\mu_2}{\sigma})^2}$. Then,

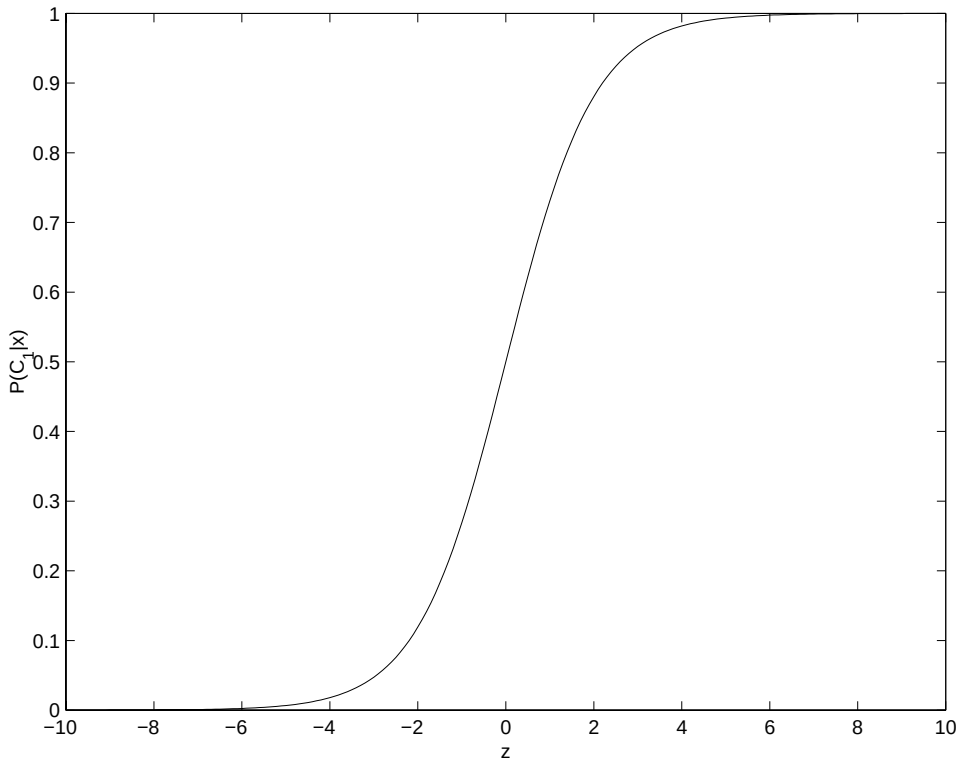

Figure 2: Logistic function

$$z = ln(\frac{P(x|C_1)P(C_1)}{P(x|C_2)P(C_2)}) = ... = (\frac{\mu_1 - \mu_2}{\sigma^2})x + ln(\frac{P(C_1}{P(C_2}) + \frac{\mu_1^2 + \mu_2^2}{2\sigma^2}$$

Let $\mu_1 > \mu_2$. Figure 3 shows the plot of the distributions.

Setting a threshold for classification is a trade-off between false positives and misses. If the σ's of the distributions are equal and the desision criteria is optimaly set, probability of misclassification can be given as:

$P(misclassification) = P(C_2) \int_{\frac{\mu_1+\mu_2}{2}}^{\infty} P(x|C_2) \ + P(C_1) \int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} P(x|C_1)$

i.e. it depends on the "overlap" of the distributions for the two classes.

There are two basic approaches to classify the input data: **generative** and **discriminative**.

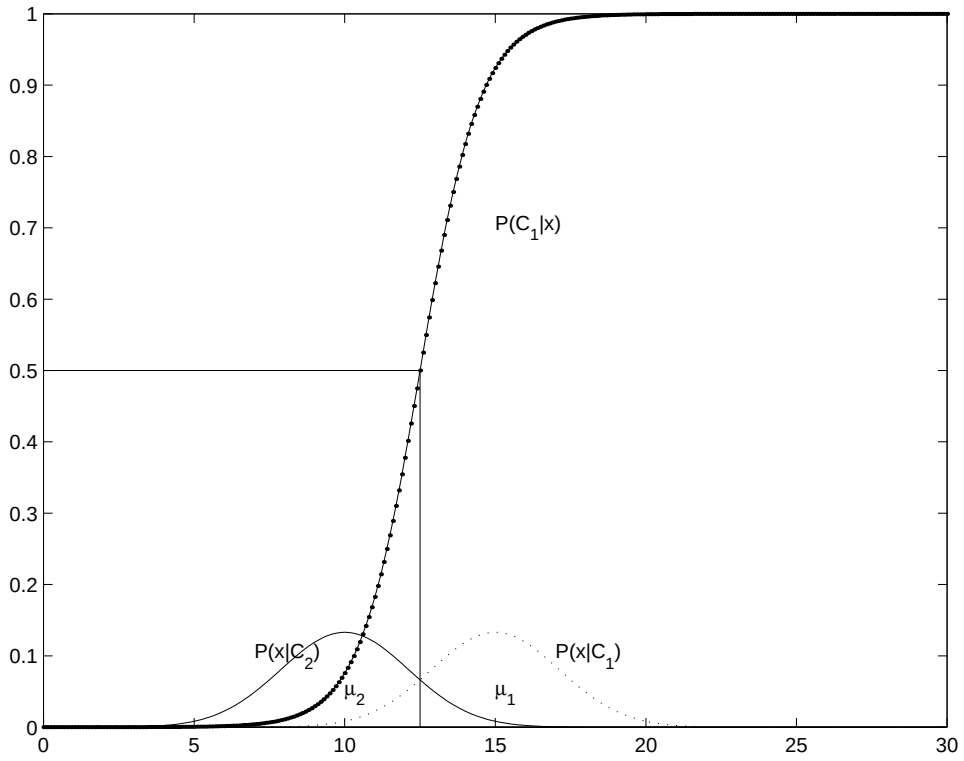

Figure 3: Probability distributions

Approach A: Generative.
• Guess suitable models for $P(x|C_1)$ and $P(x|C_2)$. Estimate parameters from the data.
• For a sample to be classified compute $P(C_1|x)$.
• Declare C_1 if $P(C_1|x) > 0.5$, and otherwise C_2.

Approach B: Discriminative.
• Learning posterior probability distributions directly.

For many distributions, including but not limited to Gaussian, we can write $z = w_1 x + w_2$ and $P(C_1|x) = \frac{1}{1+e^{-z}}$

A single layer perceptron neural network can be build to estimate the

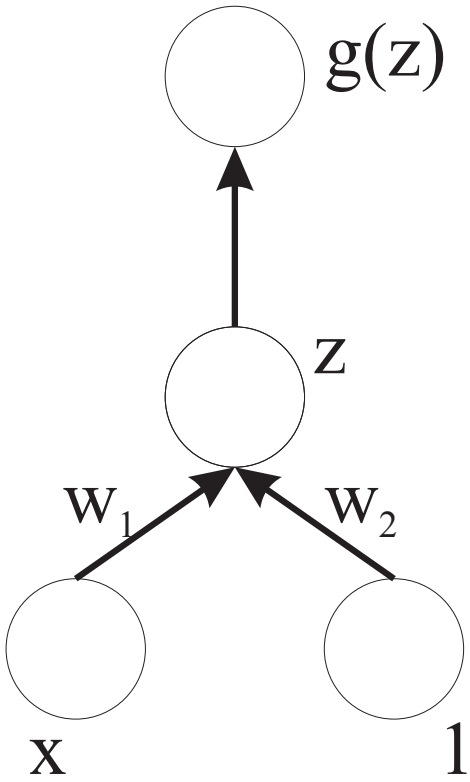

Figure 4: Single layer perceptron

weights w_1 and w_2 (figure 4) from training examples $\{x_i, (0,1)_i\}$. This is called supervised learning.

Weights can be estimated using gradient descent for SLP. Alternatively, nearest neighbor approach can be used to classify the data using appropriate metric for the input space.