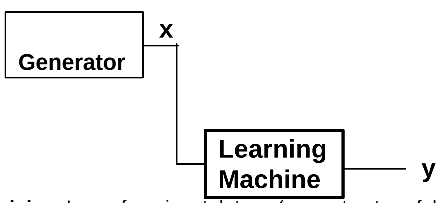
31YB: Lecture 10
Monday, 3 Nov

Overview

- Review Unsupervised Learning
- Example of Unsupervised Learning as Probability Density Estimation
- Introduction to Radial Basis Function (RBF) Network
- *Reading:* Beale and Jackson, pp. 94-96, 31YB Handout 3

Recap: Unsupervised Learning



Training: Learn from input data, **x** (e.g. extract useful information from data and form a new representation of the data **x** (see Handout 2 for more details)
Testing: Given any **x**, output a value **y** ‘describing’ **x** (such as in problems of prototyping, clustering similarity - see Handout 2 for more details)

Example of Unsupervised Learning:
Probability Density Function (PDF) Estimation
(*NOT Examinable*)

The PDF p(x) describes the distribution of x. The probability of x lying in some interval is obtained by integrating p(x) over that interval:

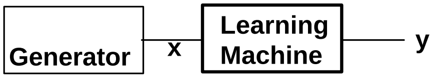
$$\mathrm{P}(a < x \le b) = \int_a^b \mathrm{p}(x)\,dx$$

$$\mathrm{P}(-\infty < x \le \infty) = \int_{-\infty}^{\infty} \mathrm{p}(x)\,dx = 1$$

Careful definition: $\mathrm{p}(a) = \lim_{e \to 0} \left[\frac{\mathrm{P}(a < x \le a + e)}{e} \right]$

Example of Unsupervised Learning:

Probability Density Function (PDF) Estimation



Training: Estimate PDF from input data, **x**
Testing: Given any **x**, output **y** = p(**x**)

Aside: Review Probability (NOT Examinable)

P(A) denotes the probability of A

$$0 \le P \le 1$$

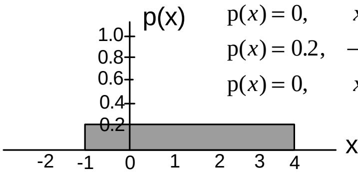
$$\mathrm{P}(A) = \lim_{n \to \infty} \frac{\text{number of ocurrences of A}}{n}$$

$$\mathrm{P}(A) = \text{degree of belief in } A$$

Example PDFs:

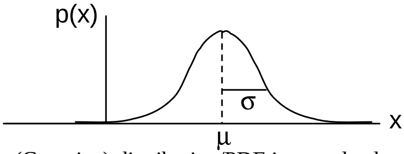
Uniform distribution (NOT Examinable)

x uniformly distributed between -1 and +4



$$\mathrm{p}(x) = 0, \quad x < -1$$
$$\mathrm{p}(x) = 0.2, \quad -1 \le x < 4$$
$$\mathrm{p}(x) = 0, \quad x \ge 4$$

Example PDFs: Gaussian (normal) distribution



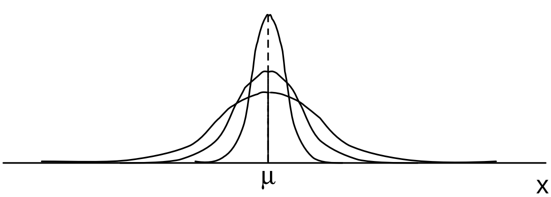
Above (Gaussian) distribution/PDF is completely characterised by its following TWO parameters:

μ is called the 'mean' (the expected value)

σ^2 is the 'variance' (it controls the spread)

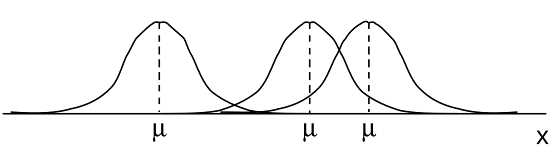
(σ is 'standard deviation')

Gaussian PDF: Effect of the σ^2 parameter



Decreasing variance (σ^2) means decreasing uncertainty about the value of x

Gaussian PDF: Effect of the μ parameter



Changing the mean (μ) changes the expected value of x

Example: 1-D Gaussian PDF

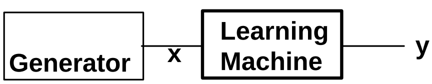
(NOT Examinable)

$$p(x) = \frac{1}{s\sqrt{2p}} \exp\left\{ -\frac{(x-m)^2}{2s^2} \right\}$$

This part ensures that p(x) is a PDF i.e. $\int_{-\infty}^{\infty} p(x)dx = 1$

An exponential function (maximum when x=μ)

Recap: Example of Unsupervised Learning: Probability Density Function (PDF) Estimation using ANN



Training: Estimate PDF from input data, **x**

Testing: Given any **x**, output **y** = p(**x**)

Example: *Parametric* Density Estimation:

• Assume that the PDF has some general form, i.e. that it belongs to a certain 'family' of functions, such as Gaussian!

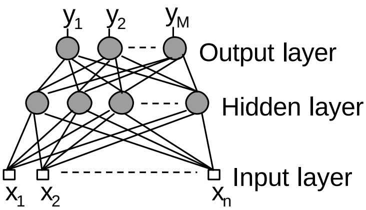
• Estimate parameters from the training data. These parameters determine the PDF (such as μ and σ^2 for Gaussian PDF) - this is exploited by Radial Basis Function Network - *next*

Alternative Paradigm to MLP: Networks of Basis Functions

- A hidden unit in an MLP has activation (output) given by *f* (**wx**) where *f* is a non-linear, **sigmoid** function

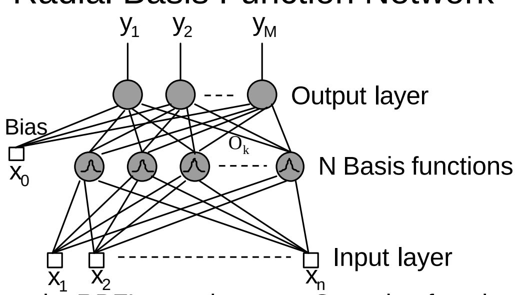
✦ Today we look at an alternative type of network in which a unit's activation (output) depends on the *distance* between its input and a stored *prototype* (i.e. *f* (**x-μ**) rather than the weighted input *f* (**wx**) as in the MLP) *– see later*

First, lets Review Feed-forward Nets (e.g. 2-layer MLP shown below)



(Note: biases not drawn for simplicity)

Radial Basis Function Network



Here the RBF's are shown as Gaussian functions.

P.B. There are NO weights connecting Input Layer and Hidden layer (comprising basis functions) – see next slide

Gaussian basis functions: *(NOT examinable)*

Each RBF hidden unit output, o_k , for k=1,…,N, is given by:

$$o_k = f(\mathbf{x}-\boldsymbol{\mu}_k) = \exp\left\{-\frac{\|\mathbf{x}-\boldsymbol{\mu}_k\|^2}{2s_k^{2}}\right\}$$

✦ Each hidden unit has parameters (m_κ, σ_κ^2); m is called the mean or expected value of Gaussian distribution , and σ^2 is called the variance (it controls the spread of the distribution about m) – see next slide example

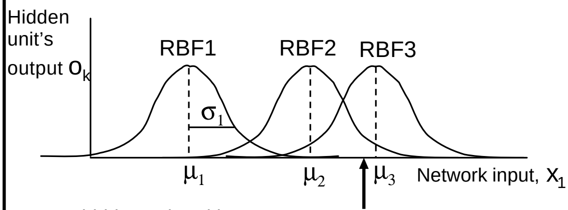
✦ Each hidden unit is *centred* on m which can be thought of as a *prototype* - *(known/stored input)*

✦ The unit's output o_k only gets large when network input **x** is *close* to its own stored prototype m_κ (as in Gaussian distribution)- closeness is measured by e.g. Euclidean distance $\|.\|^2$, which for a 1-dimensional input, x_1 is $\|\mathbf{x}-\boldsymbol{\mu}\|^2 = (x_1 - m)^2$

✦Each RBF network output, y_j , for j=1,..,M, is given by:

$$y_j = \sum_{i=0}^{N} w_i o_i = \sum_{k=0}^{N} w_k f(\mathbf{x}-\boldsymbol{\mu}_k) = \sum_{k=0}^{N} w_k \exp\left\{-\frac{\|\mathbf{x}-\boldsymbol{\mu}_k\|^2}{2s_k^{2}}\right\}$$

Radial Basis Functions: Example



✦ 3 RBF hidden units with centres or means μ_1, μ_2 & μ_3
✦ The input value indicated by arrow above results in largest output value o_3 from RBF3 (since input value is nearest to μ_3); and $o_3 > o_2 > o_1$ (since input is farthest away from centre μ_1 of RBF1).
✦Overall RBF network output is weighted sum of the above three hidden unit outputs

Notes: (see Hand-out 3 for more details)

- Each RBF node in the hidden layer responds to input only in some subspace of the input space. When input is far away from its own center μ, (many radii (standard deviations, σ) away), then the output of that unit will be so small as to be ignorable
- Each RBF has a receptive field, that is, an area of the input space to which it responds
- **BIOLOGICAL PLAUIBILITY**:
- RBF is more BIOLOGICALLY PLAUSIBLE, since many sensory neurons respond only to some small subspace of the input space, and are silent in response to all other inputs.
- There is considerable evidence that neurons in visual cortex display RECEPTIVE FIELDS of this form: they are maximally sensitive to some specific stimulus, & their output falls off as presented stimulus moves away from this "best" stimulus.

Summary

- Tutorial 2: Thurs, 6th Nov, during NEXT lecture slot (10AM), 2A87B (available on website)
- Next lecture (11): Mon, 10th Nov: **More on RBF, Training & Comparison with MLP**
- Practical 2: Thurs, 6th Nov, 2-3PM (available on website)
- Assignment due back on 20th November!
- Practical Session for help with Assignment queries??

Next Lab (Labsheet 2)

- **Character Recognition Example using the MLP**
- MLP network designed to learn to discriminate between the different letters of the alphabet. The 7 x 5 input layer receives an 'image' of a character. The output layer has 26 units, one for each letter of the alphabet.