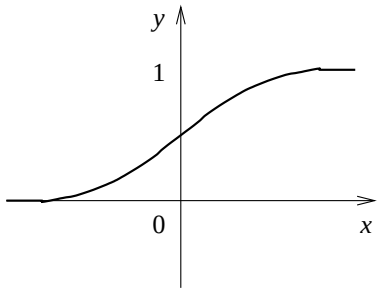
UNIVERSITY OF SURREY
Centre for Vision Speech & Signal Processing
University of Surrey, Guildford GU2 7XH.

Artificial neural networks in speech recognition

Dr Philip Jackson

- Human neural networks
  - neurons in the brain
  - inter-connectivity
- Artificial neural networks
  - types of neuron
  - network topologies
- Application to ASR
  - HMM-ANN
  - discriminative features



SIGMOID

Introduction

- Strive to match human performance
- Spoken language associated with intelligence
  - e.g., HAL in "2001: A Space Odyssey"

- **Machine intelligence**
  - rule-based systems (traditional AI, Prolog)
  - probabilistic methods (Gaussian classifiers, HMMs)
  - artificial neural networks (ANNs)

Human brain

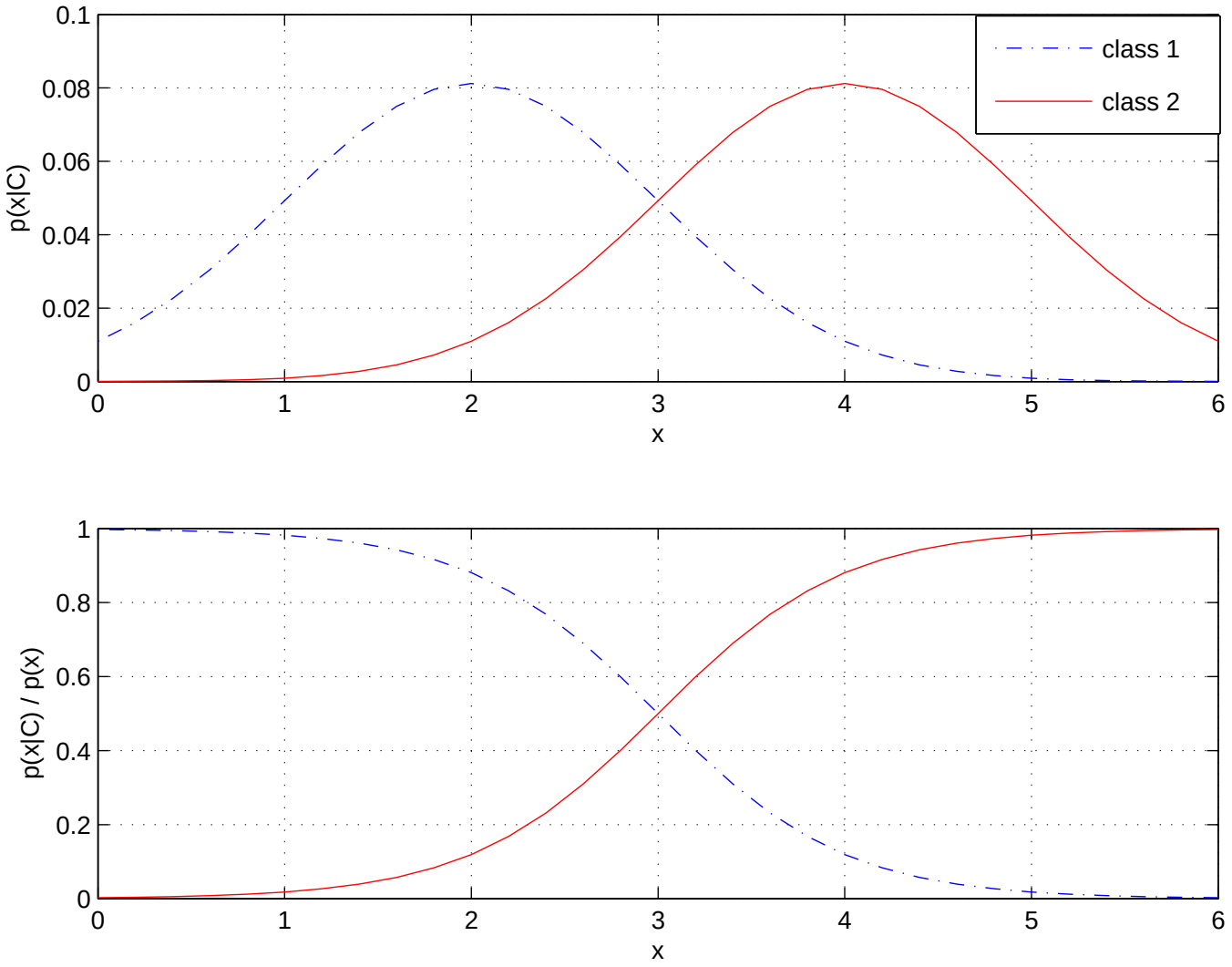
- **Anatomy of central nervous system**
  - Cerebrum has two hemispheres
  - Contains white matter and grey matter
- **Neurons and synapses**
  - Axon and dendrites
  - Connections made at synapses
  - Electro-chemical pulses at 100 m/s
  - Brain has 10^{10} neurons, 10^{13} synapses
- **Memory**
  - Long-term: learnt responses
  - Short-term: re-circulation of data

Artificial neural networks (ANNs)

- **Connectionist models:**
  - Highly inter-connected units
  - Coupling weights
  - Modified by 'learning'
  - Parallel processing

- **Types of ANN:**
  - Multi-layer perceptron (static)
  - Recurrent networks (feedback)
  - Time-delay neural networks

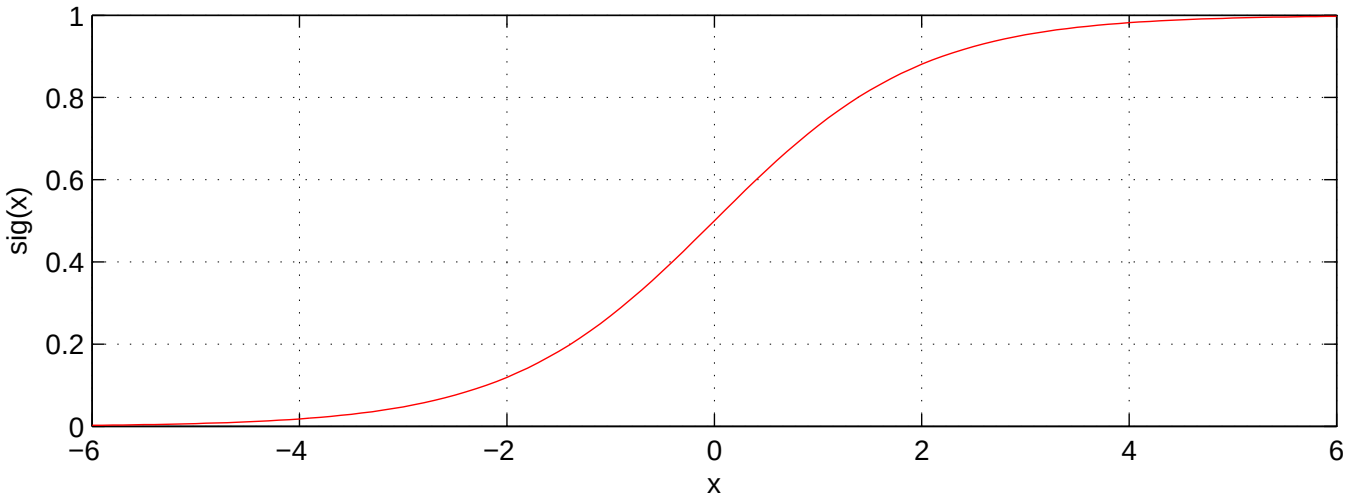
Discriminative properties of a Gaussian classifier



Classification with two Gaussian classes.

The Sigmoid function

$$f(x) \;=\; \mathrm{sig}(x) \;=\; \frac{1}{1+\exp\left(-\lambda(x-\beta)\right)} \qquad (1)$$

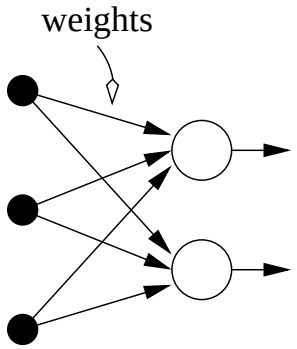


Response of a sigmoidal neuron with $\lambda = 1$, $\beta = 0$.

Properties of ANNs

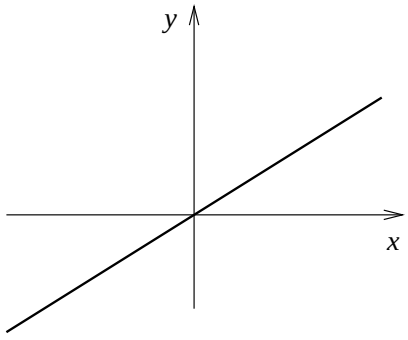
- **Single-layer perceptron**

$$x_j = \sum_i w_{ij} I_i$$

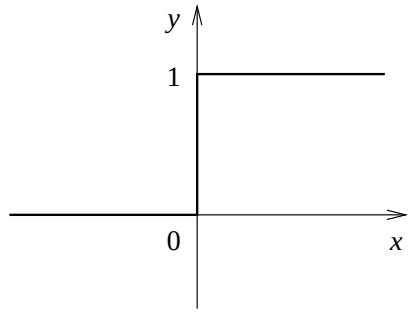


$$O_j = f\left(x_j\right)$$

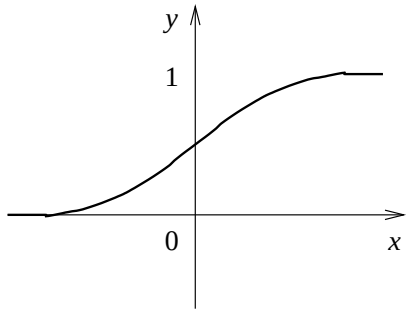
Unit responses, a.k.a. activation function



LINEAR



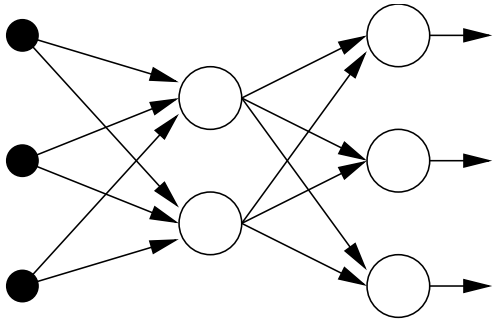
STEP



SIGMOID

Unit responses to sum of weighted inputs.

Multi-layer perceptron



$$x_j = \sum_i w_{ij}^{(1)} I_i \qquad z_k = \sum_j w_{jk}^{(2)} y_j$$

$$y_j = f^{(1)}(x_j) \qquad O_k = f^{(2)}(z_k)$$

- training by error *back propagation*

$$E_k = O_k - T_k$$

Using ANNs for speech recognition

- Innate connectivity & learnt patterns
- Difficult to interpret hidden weights

1. discriminative training (class posteriors)

2. no assumptions about form of pdfs

3. several frames provide dynamic context

4. can be easily implemented in hardware

ANNs in speech recognition

- **Successful applications:**
  - Isolated words (IWR)
  - Phoneme classification
- **Problems:**
  - segmentation needed for training
  - poor modelling of time-scale variation

Hybrid HMM-ANN methods

- Idea to combine best of HMMs and ANNs:

  1. time-domain modelling by HMM
  2. discriminative classification by ANN

- Use ANNs to compute the required emission probs for class c:

$$p(\mathcal{O}|c) = \frac{P(c|\mathcal{O})p(\mathcal{O})}{P(c)}$$

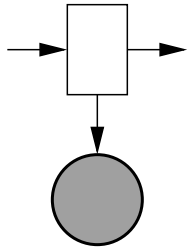
- Estimate $P(c)$ from relative freqs. in training data
- Ignore $P(\mathcal{O})$ term in comparison

- **Successful application:**

  – Large-vocabulary continuous speech recognition (LVCSR)

Comparison of HMM-GMM vs. HMM-ANN

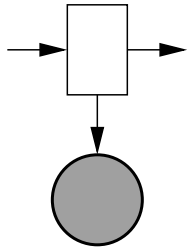
- **Conventional HMM-GMM**



Markov model, a_{ij}

GMM, $b_i(\boldsymbol{o}_t) = \sum_m c_{im} \mathcal{N}(\boldsymbol{o}_t | \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im})$

- **Hybrid HMM-ANN**



Markov model, a_{ij}

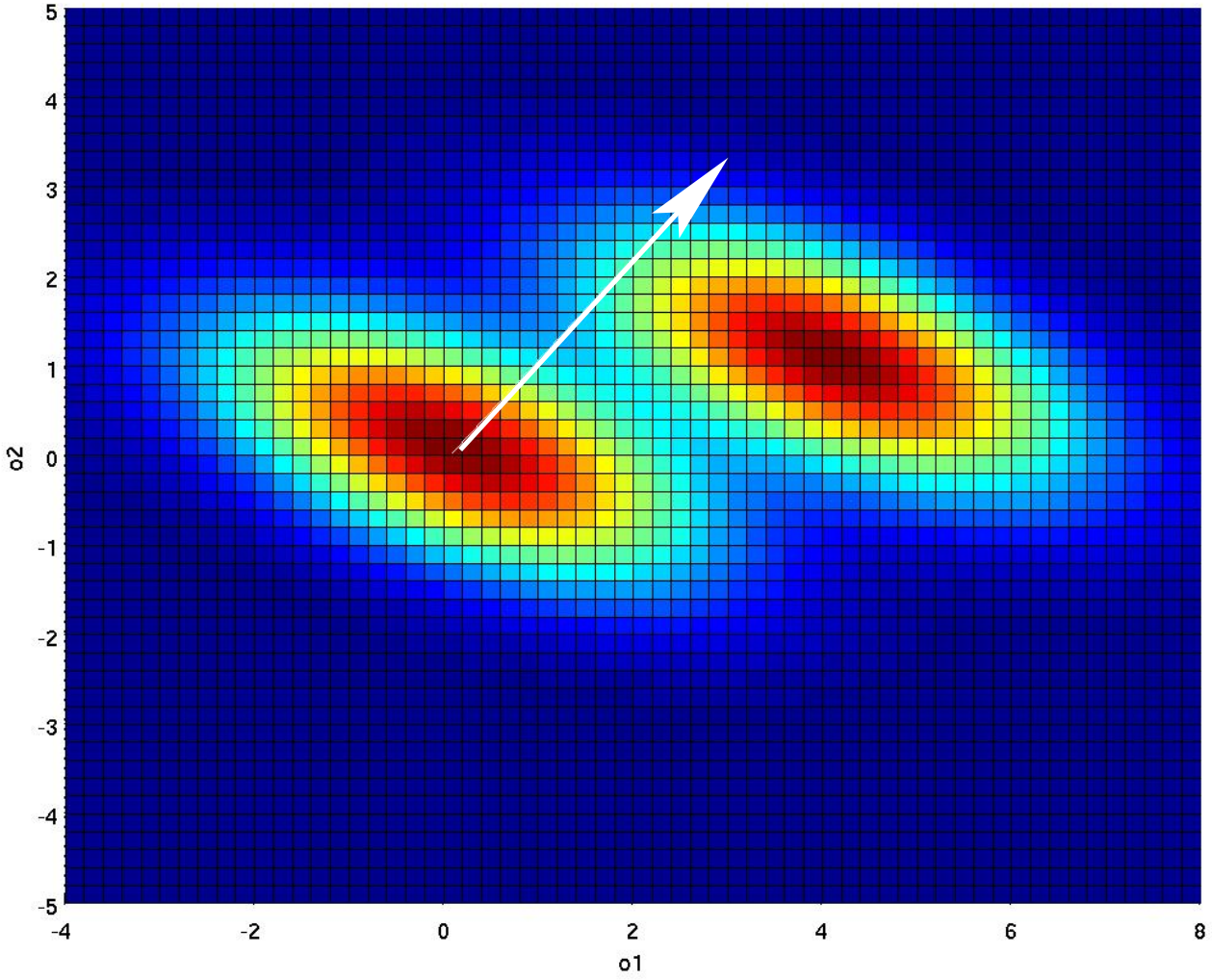
ANN, $b_i(\boldsymbol{o}_t) = P(\mathcal{O}|c)$

Feature transformations

- **Principal component analysis (PCA)**
  - PCs of X are eigenvectors of $\Sigma_X = \text{cov}(X) = \frac{XX^T}{n-1}$
  - Decomposes the correlations between features into rank-ordered modes
  - Identifies dominant dimensions of variation in data
  - Used to reduce redundancy and eliminate noise
- **Linear discriminant analysis (LDA)**
  - Maximises the ratio of between-class variance to within-class variance
  - LDs of X are eigenvectors of $\Sigma_W^{-1}\Sigma_B$
  - Decomposes the separation between classes into rank-ordered projections
  - Identifies most discriminative dimensions in the data
  - Used to ignore irrelevant or unreliable features

Discriminative features for HMMs

- First projection obtained by linear discriminant analysis



Discrimination between two Gaussians with correlation.

Summary of ANNs in ASR

- Biologically-inspired computing
- Connectionist models:
  - Multi-layered perceptron (MLP)
- Application to ASR:
  - Within HMM-ANN hybrid
  - Discriminative features

Speech recognition summary

- Introduction to ASR [**a, g1/g2, d**]
  - speech communication, source-filter theory, vocal-tract acoustics, DTW pattern matching
- Speech as spoken language [**b, j, l1/l2**]
  - phonetics, IWR/CWR/CSR grammars, phonology, morphology, syntax, language modeling
- Machine processing of speech [**e, h, m**]
  - speech spectrogram, cepstral analysis/MFCC, linear prediction/PLP, energy, delta and acceleration features
- Statistical modeling of speech [**f, i, k, n, p**]
  - MM/HMM, forward/backward procedures, Viterbi algorithm, Baum-Welch, continuous HMM, HMM-GMM
- Advanced topics in ASR [**o, q, r**]
  - context-sensitive models, MLLR/MAP adaptation, noise-robust ASR, HMM-ANN