Data Mining and Machine Learning

HMM Adaptation

Peter Jančovič

UNIVERSITY OF BIRMINGHAM

Objectives

- So far we talked about Maximum Likelihood training for HMMs (the E-M algorithm)
  – Viterbi-style training
  – Baum-Welch algorithm

- In this session, we talk about HMM adaptation:
  – Maximum A-Posteriori (MAP) estimation
  – Maximum Likelihood Linear Regression (MLLR)

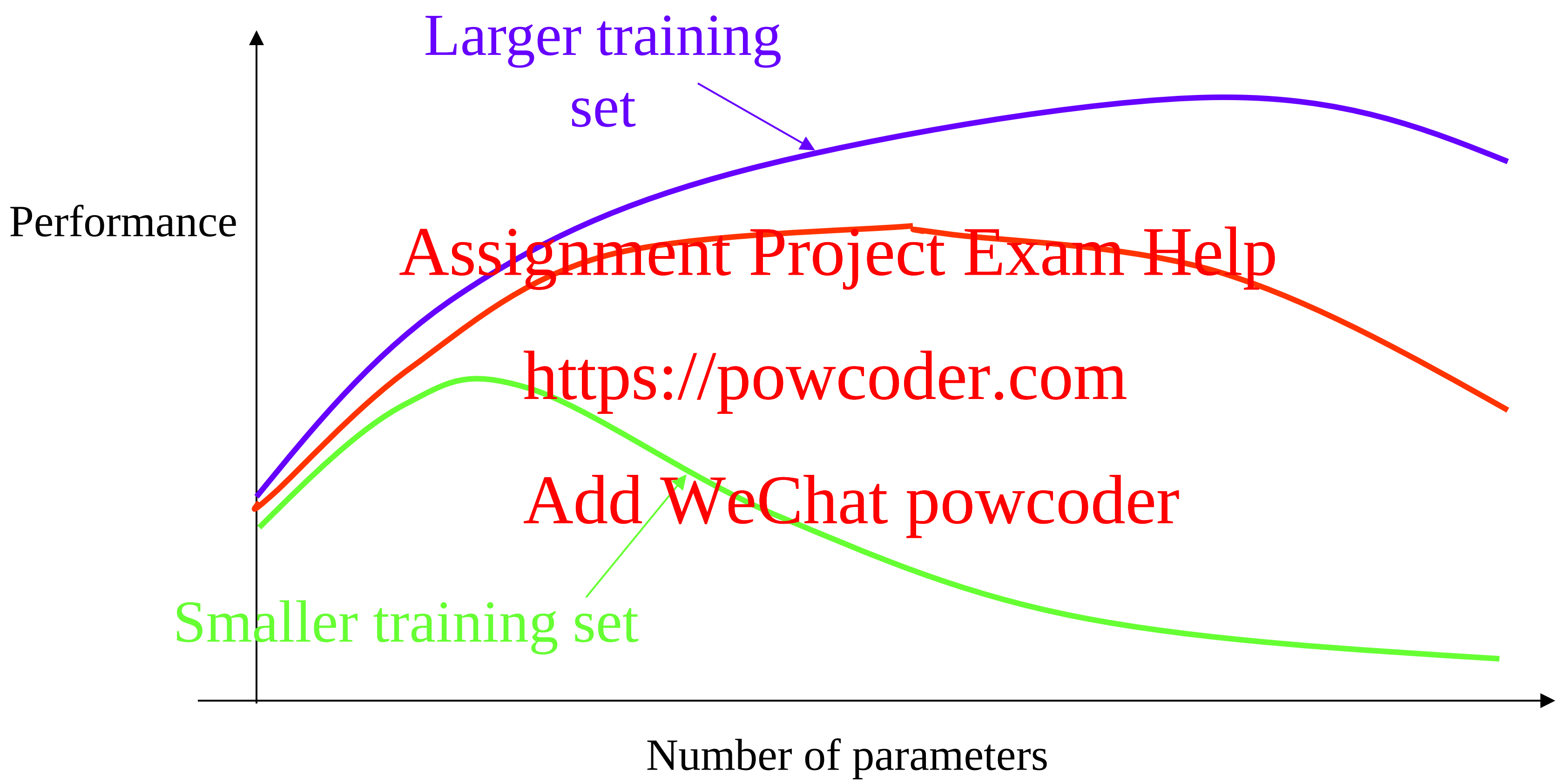
Adaptation

- A modern large-vocabulary continuous speech recognition system has many thousands of parameters
- Many hours of speech data used to train the system (e.g. 200+ hours!)
- Speech data comes from many speakers
- Hence recogniser is ‘speaker independent’
- But performance for an individual would be better if the system were speaker dependent

Adaptation

- For a single speaker, only a small amount of training data is available
- Viterbi reestimation or Baum-Welch reestimation will not work
- Adaptation:
  - the problem of robustly adapting a large number of model parameters using a small amount of training data

‘Parameters vs training data’



UNIVERSITY OF BIRMINGHAM

Adaptation

- Two common approaches to adaptation (with small amounts of training data)
  - Bayesian adaptation (also known as MAP adaptation (MAP = Maximum a Posteriori))
  - Transform-based adaptation (also known as MLLR (MLLR = Maximum Likelihood Linear Regression))

Bayesian (MAP) adaptation

- MAP estimation maximises the posterior probability of *M* given the data *y*, i.e., $P(M \mid y)$
- From Bayes' Theorem:

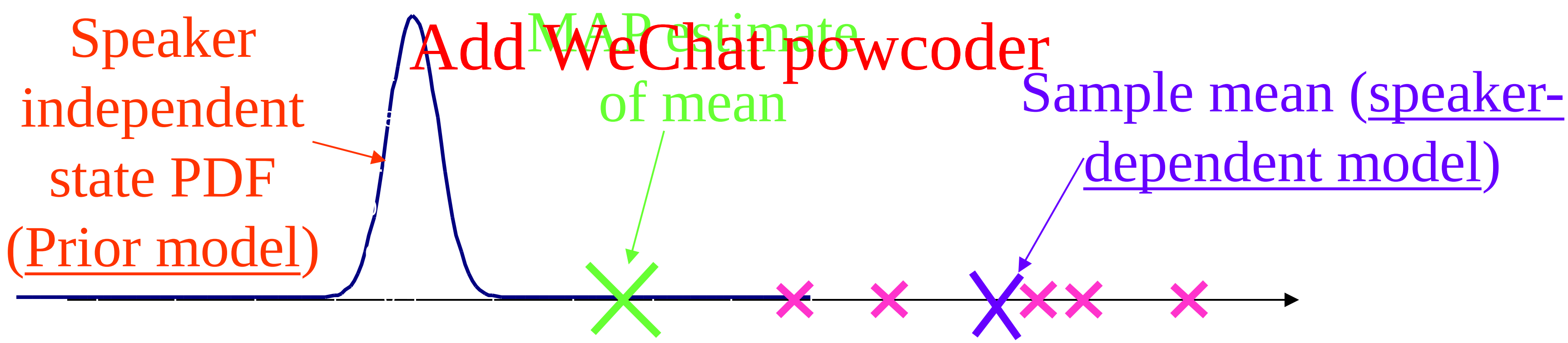
$$P(M \mid y) = \frac{p(y \mid M)P(M)}{p(y)}$$

- $P(M)$ is the prior probability of *M*
- $p(y \mid M)$ is the likelihood of the adaptation data on *M*

Bayesian (MAP) adaptation

- Uses well-trained, ‘speaker-independent’ HMM as a prior *P(M)* for the estimate of the parameters of the speaker dependent HMM

- E.G:



Bayesian (MAP) adaptation

$$\hat{M} = \lambda M_{prior} + (1-\lambda) M_{y}, 0 \le \lambda \le 1$$

MAP model

Prior model

'Speaker-dependent' model

- Intuitively, if the adaptation data set *y* is big, then the MAP adapted model will be biased towards *y*, so λ will be small
- Conversely, if there is very little adaptation data, the MAP model will be biased towards the prior, so λ will be big

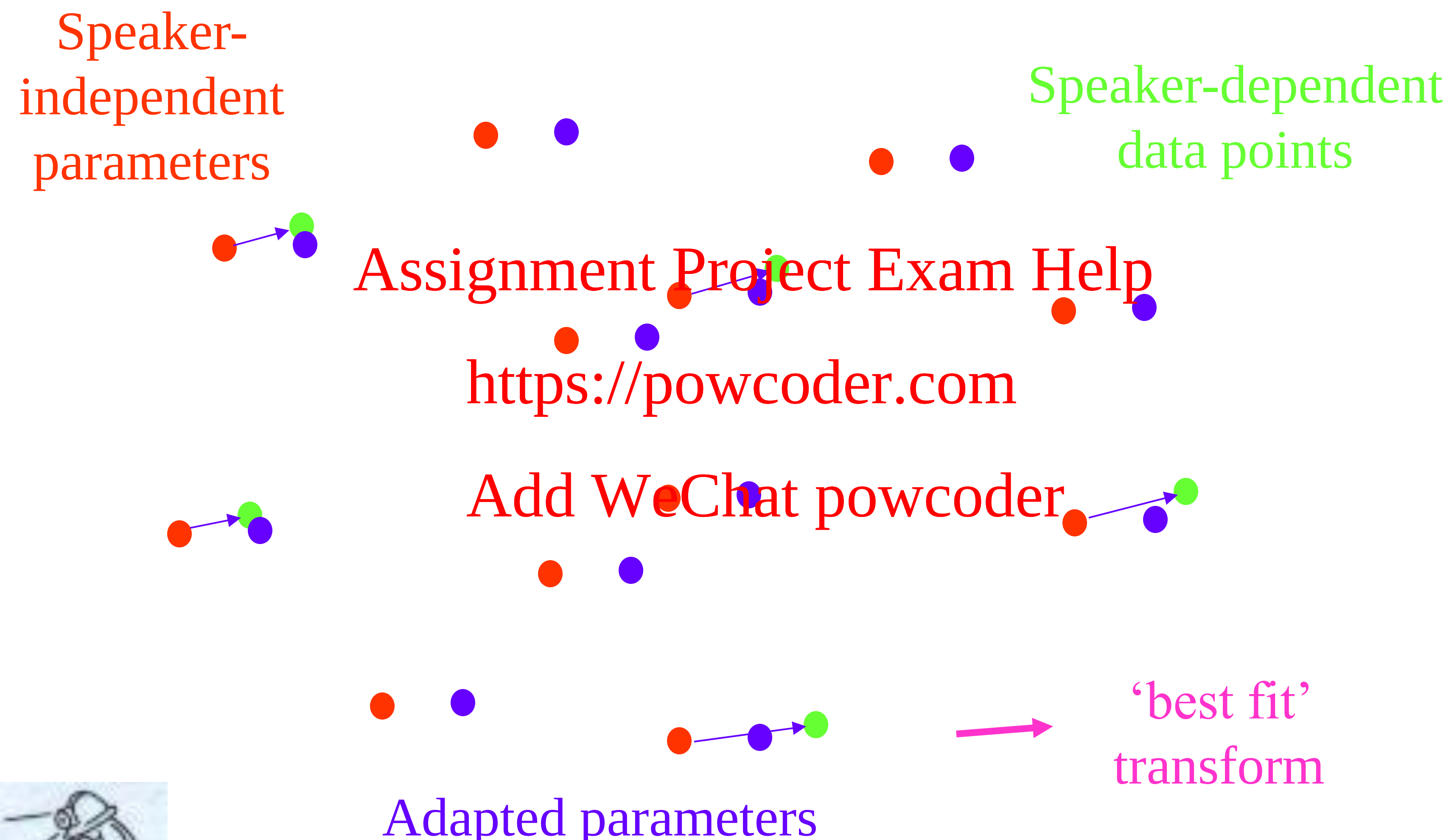
Transform-based adaptation (MLLR)

- Maximum Likelihood Linear Regression (MLLR) is another method for adapting the mean vectors of a set of HMMs
- Estimate a linear transform to transform speaker-independent into speaker-dependent parameters
- Suppose that M_{SI} is a speaker-independent HMM with Gaussian Mixture state output PDFs
- Suppose A is linear transformation on the D-dimensional space of acoustic vectors and that b is an acoustic vector
- Let $M_{SD} = T(M_{SI})$ be the HMM derived from M_{SI} by replacing each Gaussian mean vector μ with $A\mu + b$

MLLR adaptation

- Given data y from a new speaker, the aim of MLLR is to find A and b such that $P(y|T(M_{SI}))$ is maximised
- … hence Maximum Likelihood LR
- Need to estimate the $D \times D$ parameters of A
- Each acoustic vector is typically 40 dimensional, so a linear transform of the acoustic data needs 40*40 = 1600 parameters
- This is much less than the 10s or 100s of thousands of parameters needed to train the whole system
- Same transformation A can be used for all models and states.
- Alternatively, if there is enough data from the new speaker, a separate transformation can be estimated for each model, state, or set of states

Transform-based adaptation



Summary

- Bayesian (MAP) adaptation
  - J-L Gauvain and C-H Lee, “Bayesian learning for Hidden Markov Models with Gaussian mixture state observation densities”, *Speech Communication* 11, pp 205-213, 1992
- Transform-based (MLLR) adaptation
  - C J Leggeter and P C Woodland, “Maximum likelihood linear regression for speaker adaptation of continuous density HMMs”, *Computer Speech and Language*, 9, pp 171-186, 1995

UNIVERSITY OF BIRMINGHAM