More on Hidden Markov Models and their Applications

Lecture by Margrit Betke

Reading: Rabiner’89, Vogler’98

Four Classes of HMM Outputs

$b_j(0)$	Scalar Output	Vector Output
Symbols: Discrete Event Space	e.g., word b_j ("baby")	e.g., weather info b_j ([sunny, windy])
Numerical Measurements: Continuous Event Space	e.g., temperature b_j (60 F)	e.g., 3D position b_j ([x,y,z])

Last class: Discrete Scalar Output

$b_j(0)$	Scalar Output	Vector Output
Symbols: **Discrete Event Space**	e.g., word b_j ("baby")	e.g., weather info b_j ([sunny, windy])
Numerical Measurements: **Continuous Event Space**	e.g., temperature b_j (60 F)	e.g., 3D position b_j ([x,y,z])

Today: Continuous Output

$b_j(0)$	**Scalar Output**	**Vector Output**
Symbols: **Discrete Event Space**	e.g., word b_j (“baby”)	e.g., weather info b_j ([sunny, windy])
Numerical Measurements: **Continuous Event Space**	e.g., temperature b_j (60 F)	e.g., 3D position b_j ([x,y,z])

Today: Continuous Output

$b_j(0)$	Scalar Output	Vector Output
Symbols: Discrete Event Space	e.g., word b_j (“baby”)	e.g., weather info b_j ([sunny, windy])
Numerical Measurements: Continuous Event Space	e.g., temperature b_j (60 F) Density Function, e.g., Normal/Gaussian: *N*(rand. variable,mean,variance)	e.g., 3D position b_j ([x,y,z]) *N*(rand. variable, mean, variance matrix)

Mini-Intro to Estimation

Why needed?

We need to estimate the output probabilities when we train a hidden Markov model.

Scalar world: Given x_1 , …, x_n measurements (= samples)

Average $\bar{x} = 1/T\ (x_1 + \dots + x_n)$ called "sample mean"

Sample variance: $s^2 = 1/T\ \Sigma_{t=1 \text{ to } n} (x_t - \bar{x})^2$

$(\bar{x}, s^2)$ are good estimates for (μ, σ^2) of a normal density function (Gaussian) N

Mini-intro to Estimation

Vector World:

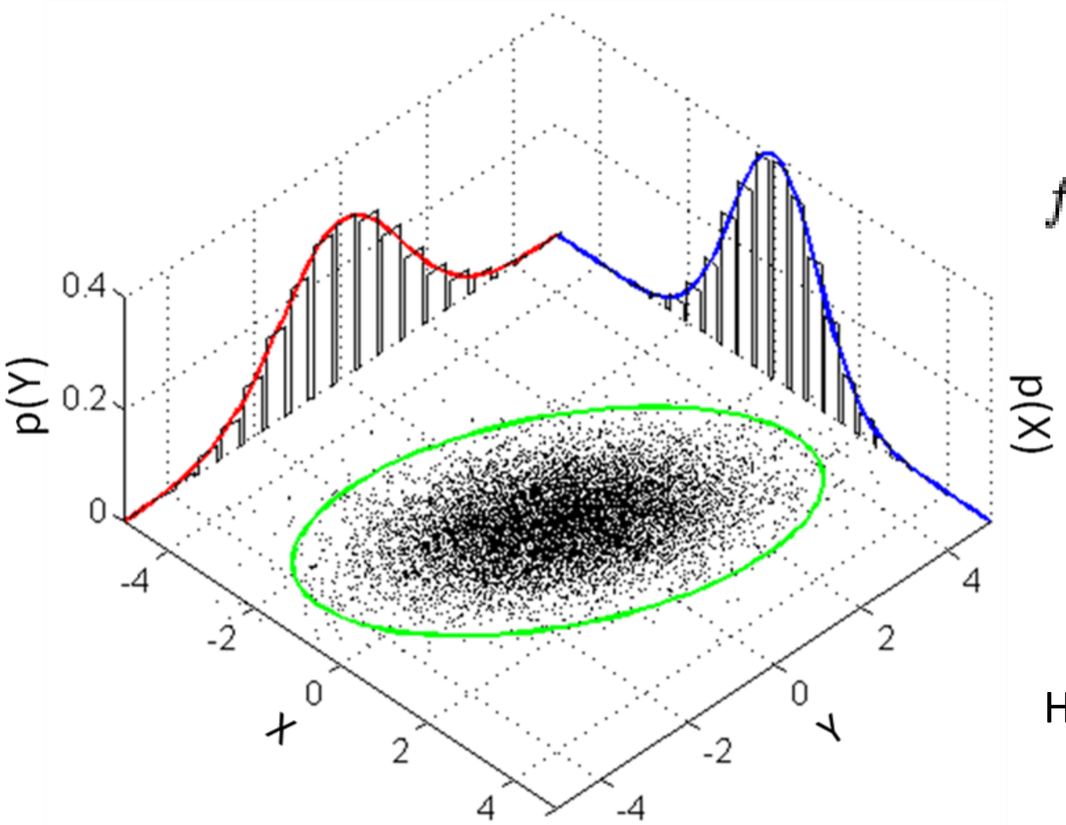
Sample mean: $\bar{\mathbf{x}} = 1/T\ (\mathbf{x}_1 + \ldots + \mathbf{x}_n)$

Sample variance: $s^2 = 1/T\ \Sigma_{t=1 \text{ to } n} (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})^T$

$(\bar{\mathbf{x}}, s^2)$ are good estimates for (μ, U) of a normal density function (Gaussian) N

Bivariate Gaussian

- https://en.wikipedia.org/wiki/Multivariate_normal_distribution



$$f_{\mathbf{X}}(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

ρ = correlation coefficient

Here: Mean $\boldsymbol{\mu} = (0,0)^T$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 3/5 \\ 3/5 & 2 \end{bmatrix}$$

Image credit: User Bscan, Wikipedia

Vogler & Metaxas

- 53 sign vocabulary, e.g., college, friend, name, I, you , what, why, have, give, win, deaf, happy, if, for
- Features: 3D wrist position, wrist orientation, velocities

Isolated Word recognition with HMMs:

10,000 experiments, ¾ training, ¼ testing per sign (178 examples)

Using 3D wrist position only: 98.4% (+/- 1%) mean performance

Adding wrist orientation: 98.2% (+/- 0.1%)

Using just velocities: 96.9% (+/-1.2%)

Vogler & Metaxas

Continuous ASL Recognition:

- 486 ASL sentences: 389 training, 97 testing
- Recognition rate: 87%

Left-to-right HMMs: Transitions occur only left to right, never backward

Output probabilities b:

Single Gaussians (= no mixtures) with diagonal covariance matrices

Vogler & Metaxas

Difficulties:

Feature selection: Variability, reliability, information content

Intra and inter signer variability (e.g., length of sign)

Gaussian densities sometimes not good model

Speed up of Recognition:

Add “Beam searching” to Viterbi Algorithm:

Threshold on $d_t(i)$. If too low, partial path probability too low. Probably does not contribute to most likely path

-> Set to zero.