CS485/685
Lecture 14: Feb 16, 2012

HMMs continued

[B] Sections 13.1-13.2

Case study: actigraphy

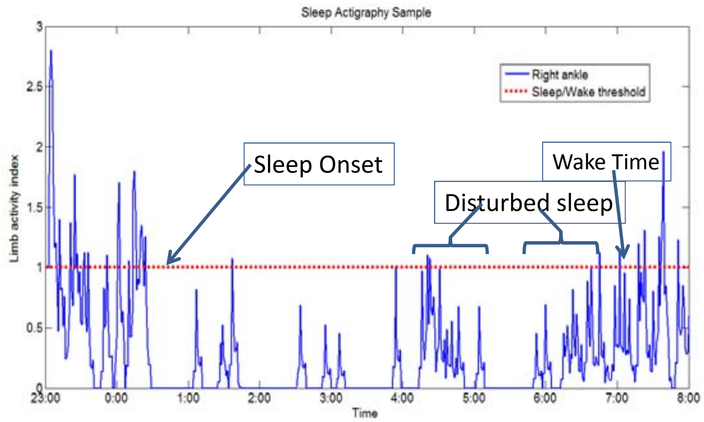
- Task: infer periods of rest vs activity from accelerometer data

ankle

wrist

Sleep Detection

- Simple thresholding approach



Alternative: HMM model

- Segment data into short periods of time (e.g., 1 min)
- Monitoring: probability of sleep in each segment

Parameterization

- Class: Y
- Measurement: X
  - Features: $\phi(X)$
- Initial distribution: $\Pr(Y_0)$
- Transition dist: $\Pr(Y_t|Y_{t-1})$
- Emission dist: $\Pr(X_t|Y_t)$

Monitoring

- Suppose we observe the following sequence of features: $X_{1..3} = (3.5, 0.5, 0.7)$
- What is the probability of $Y_t = sleep$ at each time step?
- **Forward algorithm**: iterate
  $\Pr(y_i|x_{1..i}) \propto \Pr(x_i|y_i) \sum_{y_{i-1}} \Pr(y_i|y_{i-1}) \Pr(y_{i-1}|x_{1..i-1})$

Example

Most likely explanation

- In actigraphy, we are not interested in estimating the probability of sleep at each time step in isolation
- Instead, we want the most likely explanation (i.e., sequence of classes) of the measurements

$$argmax_{y_0,\ldots,y_t} \Pr(y_{0..t}|x_{1..t})$$

- **Viterbi algorithm:** iterate

$$\max_{y_{0..i}} \Pr(y_{0..i+1}|x_{0..i}) \propto$$
$$\max_{y_i} \Pr(y_{i+1}|y_i) \Pr(x_i|y_i) \max_{y_{0..i-1}} \Pr(y_{0..i}|x_{0..i-1})$$

Example

Parameter Learning

- Parameters:
  - Initial distribution: $\Pr(Y_0 = y) = p_y$
  - Transition distribution: $\Pr(Y_i = y|Y_{i-1} = y) = p_{y|y}$
    $\Pr(Y_i = y|Y_{i-1} = \sim y) = p_{y|\sim y}$
  - Emission distribution:
    - Multinomial: $\Pr(X_i = x|Y_i = y) = p_{x|y}$
      $\Pr(X_i = x|Y_i = \sim y) = p_{x|\sim y}$
    - Gaussian: $\Pr(X_i = x|Y_i = y) = N(\mu_y, \sigma_y)$
      $\Pr(X_i = x|Y_i = y) = N(\mu_{\sim y}, \sigma_{\sim y})$

Maximum Likelihood

- Supervised Learning: Y's are known
- Objective: $argmax_{param} \Pr(Y_{0..t}, X_{1..t} | param)$

- Derivation:
  - Set derivative to 0
  - Isolate parameters

Multinomial emissions

- Let $\#y_0$ be # of times that process starts in y_0
- Let $\#(y', y)$ be # of times that y' follows y
- Let $\#(x, y)$ be # of times that x occurs with y
- $\Pr(Y_{0..t}, X_{1..t})$

$$= \Pr(Y_0) \prod_{i=1}^{t} \Pr(Y_i | Y_{i-1}) \Pr(X_i | Y_i)$$

$$= p_{y_0}^{\#y_0} (1 - p_{y_0})^{\#\sim y_0} p_{y|y}^{\#(y,y)} (1 - p_{y|y})^{\#(\sim y,y)}$$
$$p_{y|\sim y}^{\#(y,\sim y)} (1 - p_{y|\sim y})^{\#(\sim y,\sim y)} p_{x|y}^{\#(x,y)} (1 - p_{x|y})^{\#(\sim x,y)}$$
$$p_{x|\sim y}^{\#(x,\sim y)} (1 - p_{x|\sim y})^{\#(\sim x,\sim y)}$$

Multinomial emissions

- $argmax_{params} \Pr(y_{0..t}, x_{1..t}|params)$

$$\Longrightarrow \begin{cases} argmax_{p_{y_0}} p_{y_0}^{\#y_0}\left(1-p_{y_0}\right)^{\#\sim y_0} \\ argmax_{p_{y|y}} p_{y|y}^{\#(y,y)}\left(1-p_{y|y}\right)^{\#(\sim y,y)} \\ argmax_{p_{y|\sim y}} p_{y|\sim y}^{\#(y,\sim y)}\left(1-p_{y|\sim y}\right)^{\#(\sim y,\sim y)} \\ argmax_{p_{x|y}} p_{x|y}^{\#(x,y)}\left(1-p_{x|y}\right)^{\#(\sim x,y)} \\ argmax_{p_{x|\sim y}} p_{x|\sim y}^{\#(x,\sim y)}\left(1-p_{x|\sim y}\right)^{\#(\sim x,\sim y)} \end{cases}$$

Multinomial emissions

- Optimization problem:

$$\max_{p_a} p_a^{\#a}(1-p_a)^{\#\sim a}$$

$$\Longrightarrow \max_{p_a} \#a \log(p_a) + (\#\sim a)\log(1-p_a)$$

- Set derivative to 0:

$$0 = \frac{\#a}{p_a} - \frac{\#\sim a}{1-p_a}$$

$$\Longrightarrow (1-p_a)\#a = p_a(\#\sim a)$$

$$\Longrightarrow p_a = \frac{\#a}{\#a+\#\sim a}$$

Relative Frequency Counts

- Maximum likelihood solution

$p_{y_0} = \#y_0/(\#y_0 + \#{\sim}y_0)$

$p_{y|y} = \#(y,y)/(\#(y,y) + \#({\sim}y,y))$

$p_{y|{\sim}y} = \#(y,{\sim}y)/(\#(y,{\sim}y) + \#({\sim}y,{\sim}y))$

$p_{x|y} = \#(x,y)/(\#(x,y) + \#({\sim}x,y))$

$p_{x|{\sim}y} = \#(x,{\sim}y)/(\#(x,{\sim}y) + \#({\sim}x,{\sim}y))$

Gaussian Emissions

- Maximum likelihood solution

$p_{y_0} = \#y_0/(\#y_0 + \#{\sim}y_0)$

$p_{y|y} = \#(y,y)/(\#(y,y) + \#({\sim}y,y))$

$p_{y|{\sim}y} = \#(y,{\sim}y)/(\#(y,{\sim}y) + \#({\sim}y,{\sim}y))$

$\mu_y = \frac{1}{t}\sum_{\{i|Y_i=y\}} x_i, \sigma_y^2 = \frac{1}{t}\sum_{\{i|Y_i=y\}}\left(x_i - \mu_y\right)^2$

$\mu_{\sim y} = \frac{1}{t}\sum_{\{i|Y_i=\sim y\}} x_i, \sigma_{\sim y}^2 = \frac{1}{t}\sum_{\{i|Y_i=\sim y\}}\left(x_i - \mu_{\sim y}\right)^2$