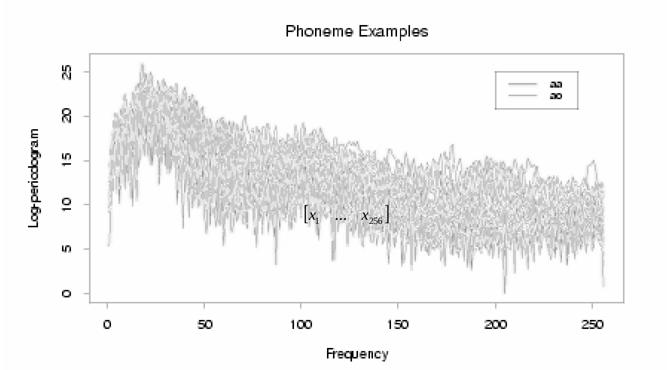
Example: Phoneme Recognition



- Plots of 15 examples of each class

$$\begin{bmatrix} x_1 & \dots & x_{256} \end{bmatrix}^T$$

Example: Phoneme Recognition

- Example $\begin{bmatrix} x_1 & \dots & x_{256} \end{bmatrix}^T$
- Observations X(f) of a grid of 256 frequencies

$$\left\{ \begin{matrix} x_{1,1} & \dots & x_{1,256} \\ \dots & & \dots \\ x_{1000,1} & \dots & x_{1000,256} \end{matrix} \right\}$$

Logistic Regression Model

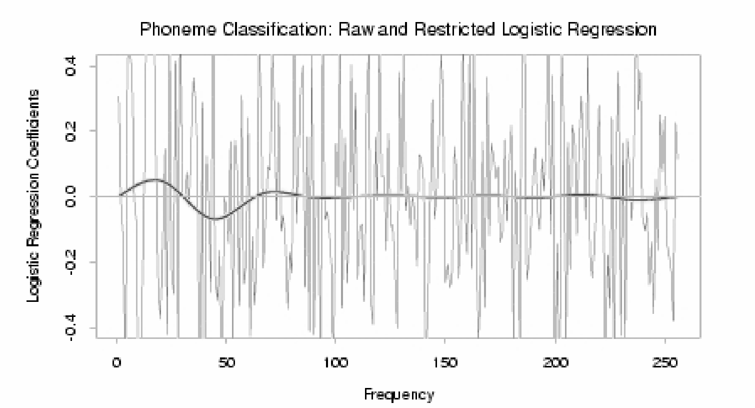
- Linear regression of $\log \frac{\Pr(aa \mid X)}{\Pr(ao \mid X)}$

$$\log \frac{\Pr(aa \mid X)}{\Pr(ao \mid X)} = \int X(f)\beta(f)df = \sum_{j=1}^{256} X(f_j)\beta(f_j) = \sum_{j=1}^{256} x_j \beta_j$$

- β_j fit by maximum likelihood
  - See http://www.stat.cmu.edu/~minka/papers/logreg.html

Problems:

- Negative autocorrelation in the β coefficients
- Only four observations per coefficient



Approximate Coefficients

- Each coefficient β_f is expressed as a sum of M=12 basis functions
- Each basis function is computed at all the data points

$$\beta_f = \underbrace{\begin{Bmatrix} h_1(x_1) & \dots & h_M(x_1) \\ \dots & & \dots \\ h_1(x_{256}) & \dots & h_M(x_{256}) \end{Bmatrix}}_{H} \underbrace{\begin{bmatrix} \theta_1 \\ \dots \\ \theta_M \end{bmatrix}}_{\theta} = \sum_{m=1}^{M=12} h_m(f)\theta_m$$

Filtered Features

- Original regression with β_f can be seen as regression with coefficients θ_f

$$\begin{bmatrix} x_1 & \dots & x_{256} \end{bmatrix}^T \begin{bmatrix} \beta_1 \\ \\ \beta_{256} \end{bmatrix} = x^T\beta = x^T H\theta = (H^t x)^T \theta = x^*\theta$$

- New features are $x^* = H^T x$

Regression with Filtered Features