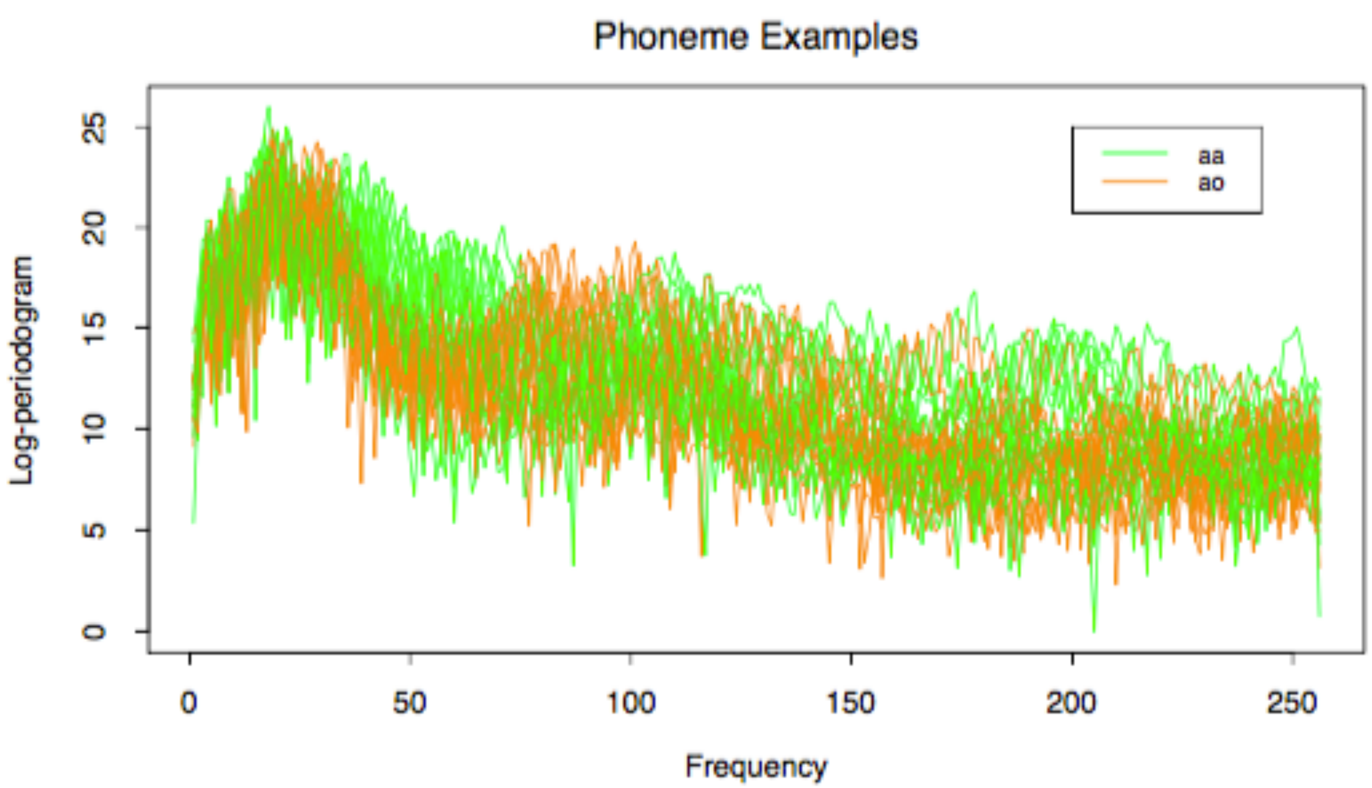
Logistic Regression with Functional Data

ESL 5.2.3 Phoneme Recognition

- Binary response Y: two classes “**aa**” (695) and “**ao**” (1022)
- Numerical feature X: log-periodogram measured at 256 uniformly spaced frequencies.
- Logistic regression model:

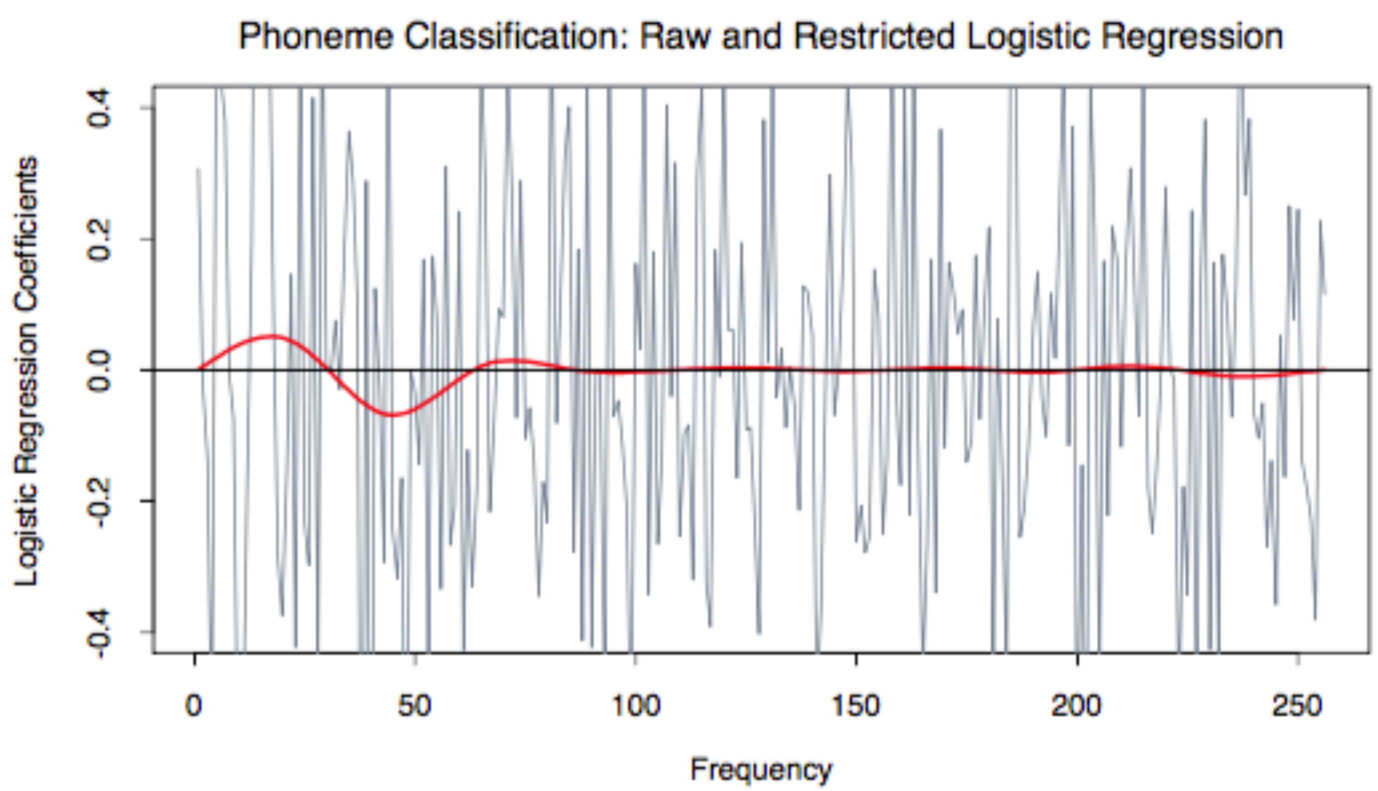
$$\boxed{\log \frac{P(aa|\mathbf{x})}{P(ao|\mathbf{x})} = \beta_0 + \sum_{j=1}^{256} x_j \beta_j}$$



ESL 5.2.3 Phoneme Recognition

- Recall the 256 measurements for each sample are not the same as measurements collected from 256 independent predictors. They are observations (at discretized frequencies) from a continuous Log-periodogram function.
- Naturally we would expect the 256 coefficients beta_j's are also continuous in the frequency domain. So we model it by splines

$$\beta(\nu) = \sum_{m=1}^{M} h_m(\nu)\alpha_m, \quad \nu = 1, 2, \ldots, 256.$$



ESL 5.2.3 Phoneme Recognition

$$\beta(\nu) = \sum_{m=1}^{M} h_m(\nu)\alpha_m, \quad \nu = 1, 2, \ldots, 256.$$

$$\mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} = \mathbf{X}_{n\times p}\mathbf{H}_{p\times M}\boldsymbol{\alpha}_{M\times 1} = \mathbf{B}_{n\times M}\boldsymbol{\alpha}_{M\times 1}$$

X to B: filtering or feature extraction

Obtain alpha: fit a logistic regression model with design matrix B

beta = H*alpha

GAM Logistic Regression

$$g((x)) = \alpha + g_1(x_1) + g_2(x_2) + \cdots + g_p(x_p)$$

$$\log \frac{p(\mathbf{x})}{1 - p(\mathbf{x})} = \alpha + g_1(x_1) + g_2(x_2) + \cdots + g_p(x_p)$$

Backfitting Algorithm

Evaluate Classification Accuracy

Confusion Matrix and ROC Curve

		Predicted Class	
		No	Yes
Observed Class	No	TN	FP
	Yes	FN	TP

TN	True Negative
FP	False Positive
FN	False Negative
TP	True Positive

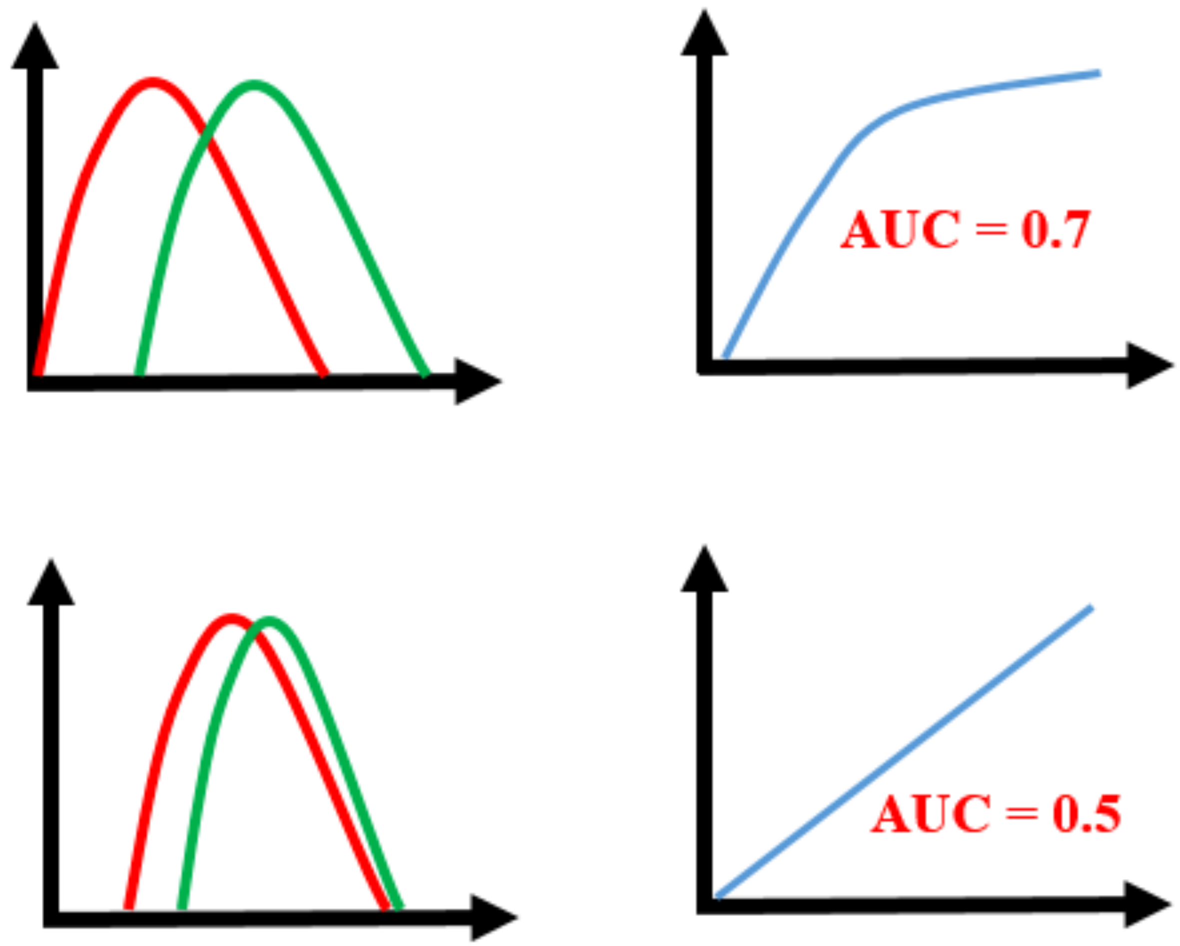
Model Performance

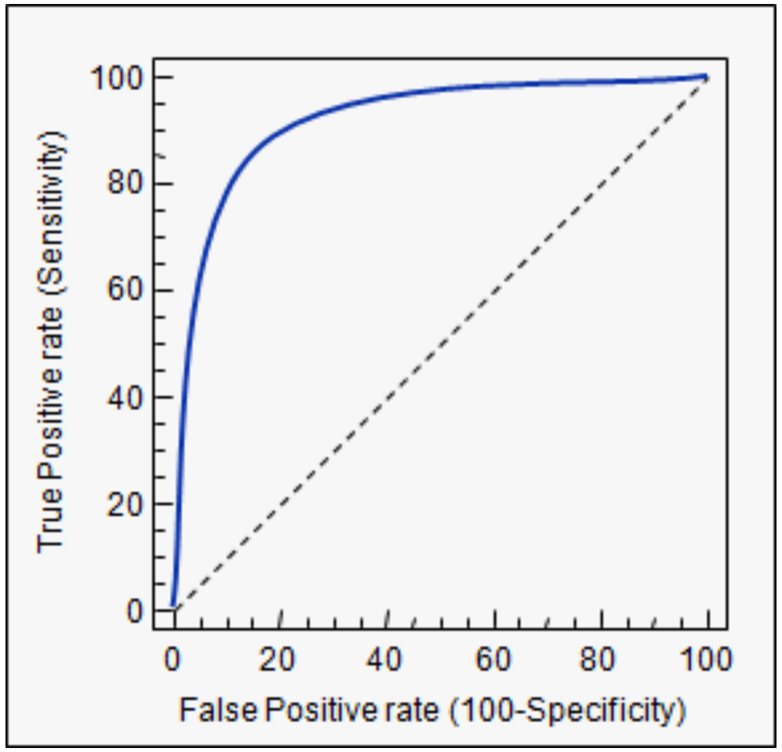
Accuracy	= (TN+TP)/(TN+FP+FN
Precision	= TP/(FP+TP)
Sensitivity	= TP/(TP+FN)
Specificity	= TN/(TN+FP)

Evaluate Classification Accuracy

		True condition				
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**, Power	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$	F_1 score = $\frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$		

AUC and ROC


AUC = 0.7
AUC = 0.5


True Positive rate (Sensitivity)
100
80
60
40
20
0
0
20
40
60
80
100
False Positive rate (100-Specificity)

AUC and Mann-Whitney U Statistic

Mann-Whitney U-stat or Wilcoxon rank sum Stat

1. Assign numeric ranks to all the observations (put the observations from both groups to one set), beginning with 1 for the smallest value. Where there are groups of tied values, assign a rank equal to the midpoint of unadjusted rankings. E.g., the ranks of $(3, 5, 5, 5, 5, 8)$ are $(1, 3.5, 3.5, 3.5, 3.5, 6)$ (the unadjusted rank would be $(1, 2, 3, 4, 5, 6)$).
2. Now, add up the ranks for the observations which came from sample 1. The sum of ranks in sample 2 is now determinate, since the sum of all the ranks equals $N(N+1)/2$ where N is the total number of observations.
3. U is then given by:[4]

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

$$\boxed{\mathrm{AUC} = U_1/(n_1 n_2)}$$