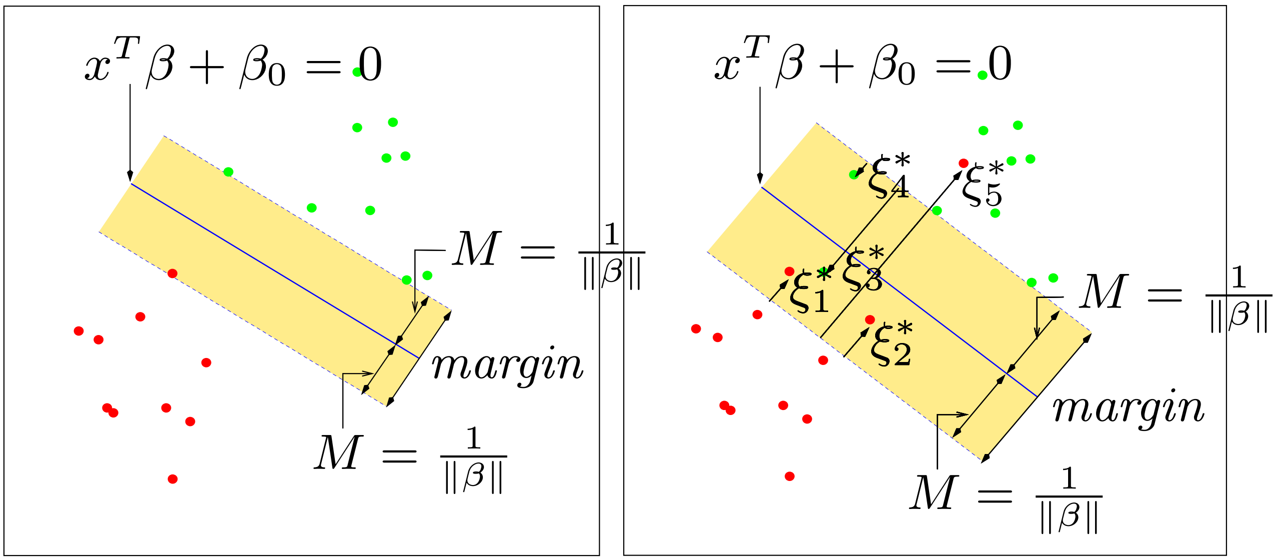
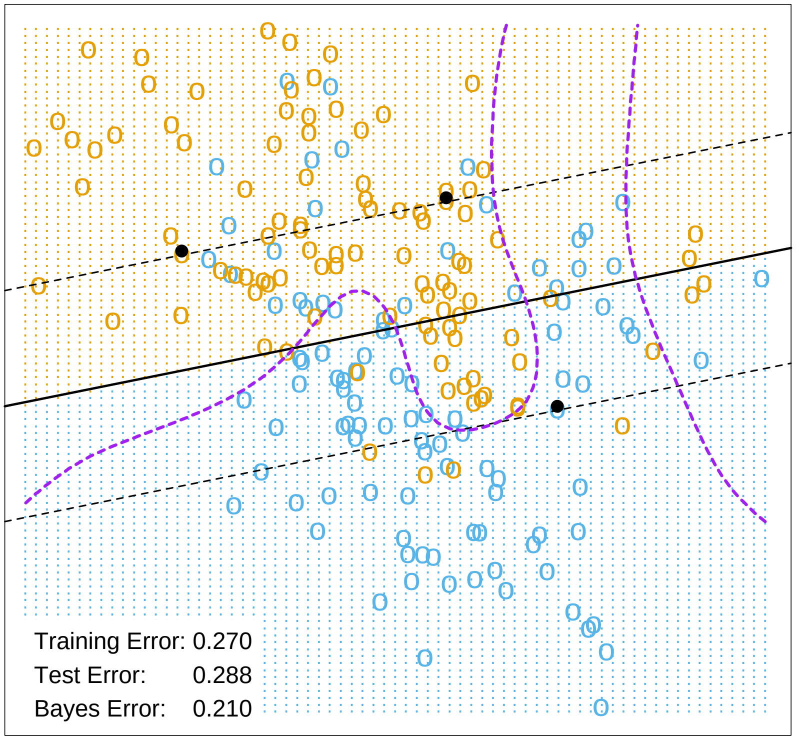
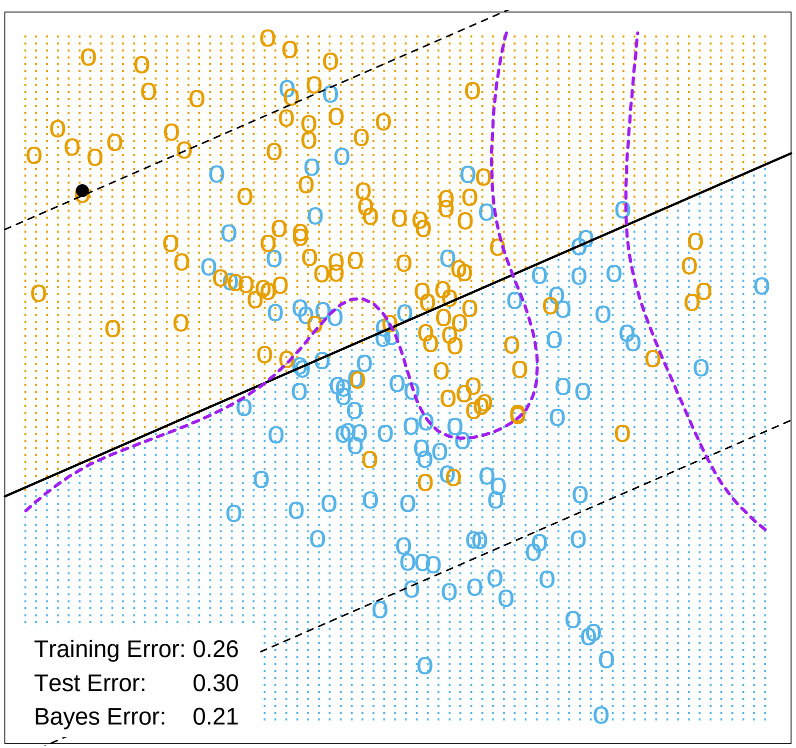



**FIGURE 12.1.** *Support vector classifiers. The left panel shows the separable case. The decision boundary is the solid line, while broken lines bound the shaded maximal margin of width* $2M = 2/\|\beta\|$*. The right panel shows the nonseparable (overlap) case. The points labeled* $\xi_j^*$ *are on the wrong side of their margin by an amount* $\xi_j^* = M\xi_j$*; points on the correct side have* $\xi_j^* = 0$*. The margin is maximized subject to a total budget* $\sum \xi_i \leq$ *constant. Hence* $\sum \xi_j^*$ *is the total distance of points on the wrong side of their margin.*




$C = 10000$




$C = 0.01$

**FIGURE 12.2.** *The linear support vector boundary for the mixture data example with two overlapping classes, for two different values of C. The broken lines*

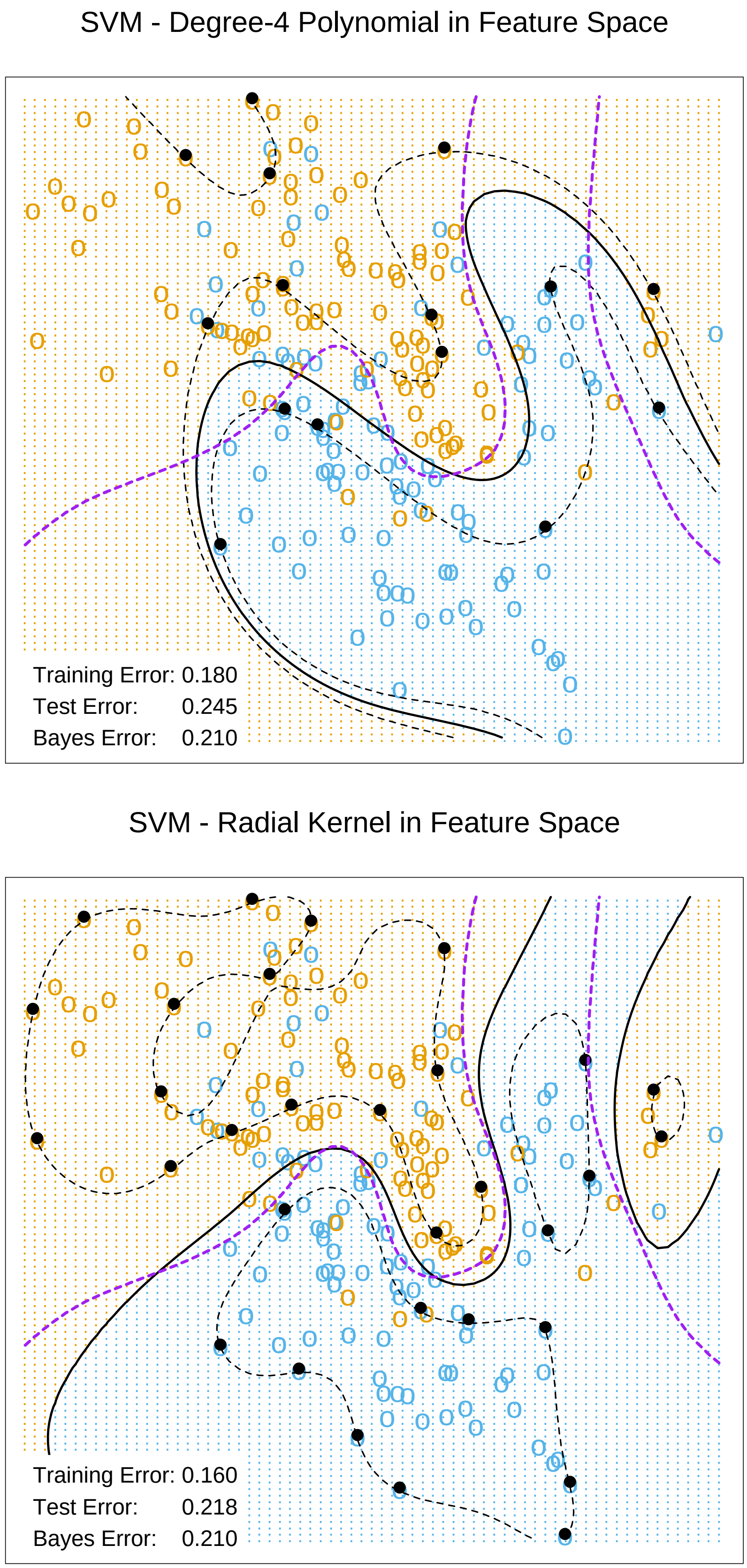



**FIGURE 12.3.** *Two nonlinear SVMs for the mixture data. The upper plot uses a 4th degree polynomial*

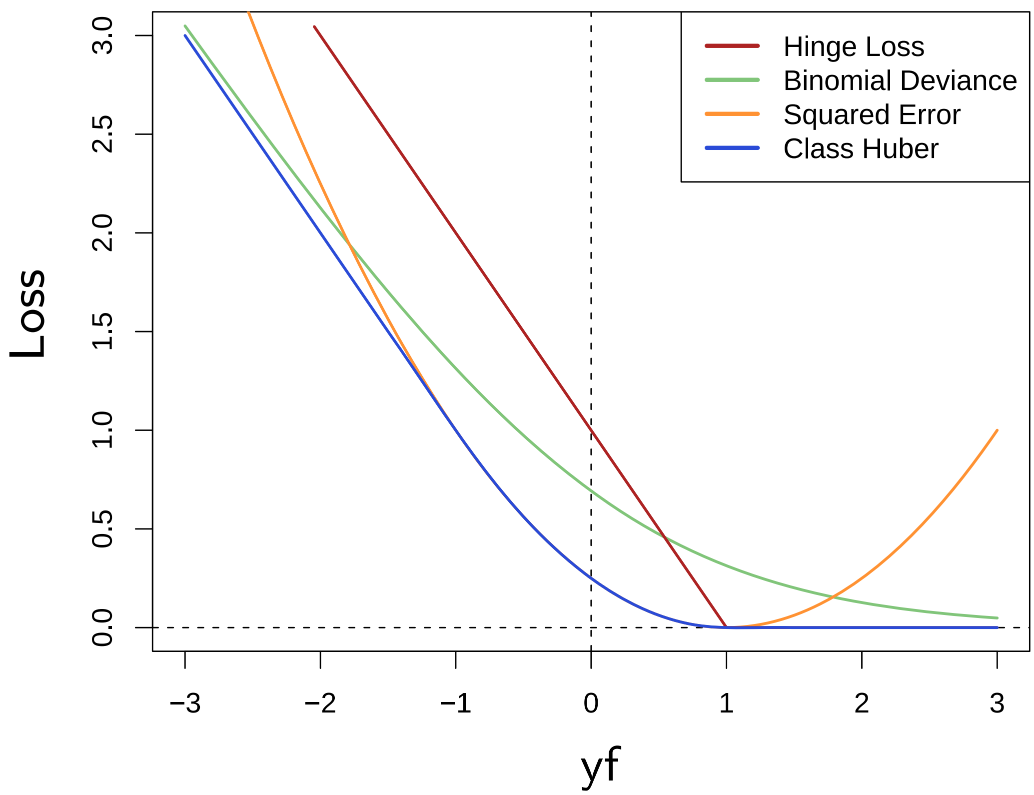
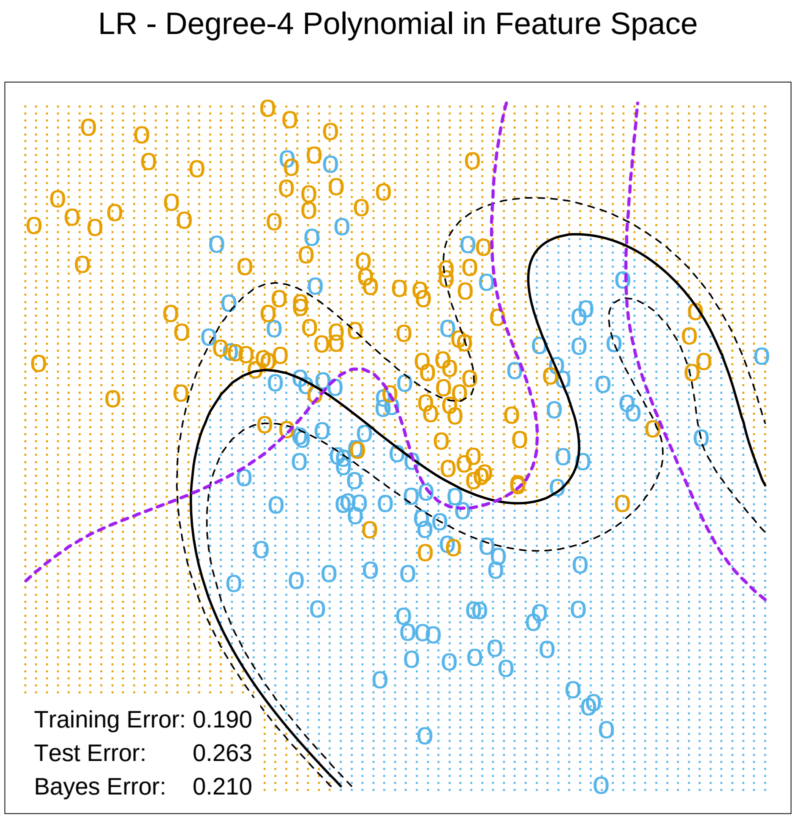



**FIGURE 12.4.** *The support vector loss function (hinge loss), compared to the negative log-likelihood loss (binomial deviance) for logistic regression, squared-error loss, and a "Huberized" version of the squared hinge loss. All are shown as a function of* $yf$ *rather than* $f$*, because of the symmetry between the* $y = +1$ *and* $y = -1$ *case. The deviance and Huber have the same asymptotes as the SVM loss, but are rounded in the interior All are scaled to have the limiting left-tail slope of* $-1$*.*



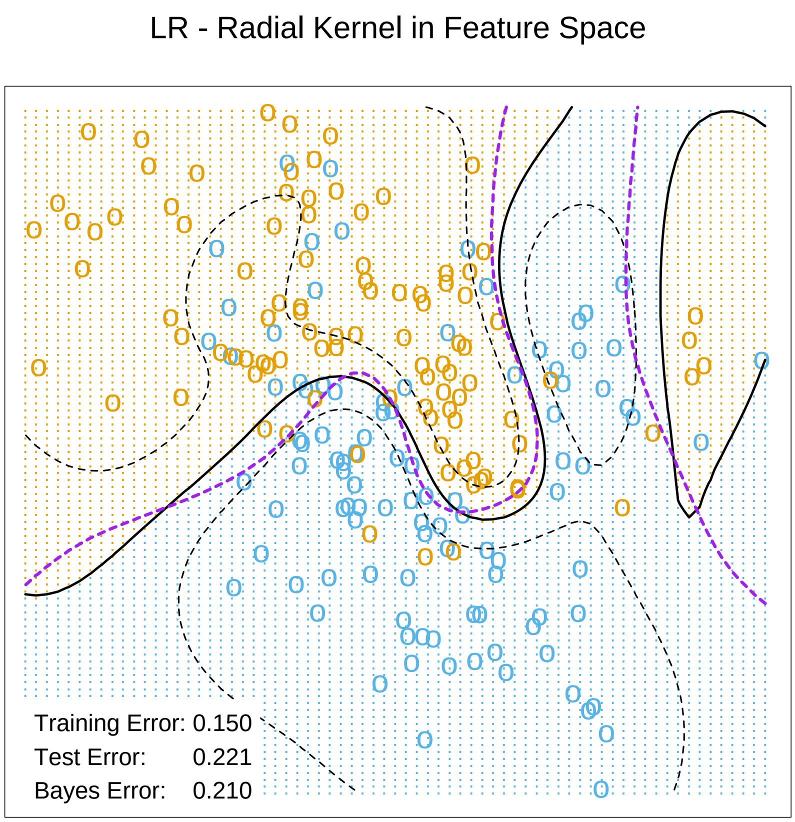




**FIGURE 12.5.** *The logistic regression versions of the SVM models in Figure 12.3, using the identical kernels*

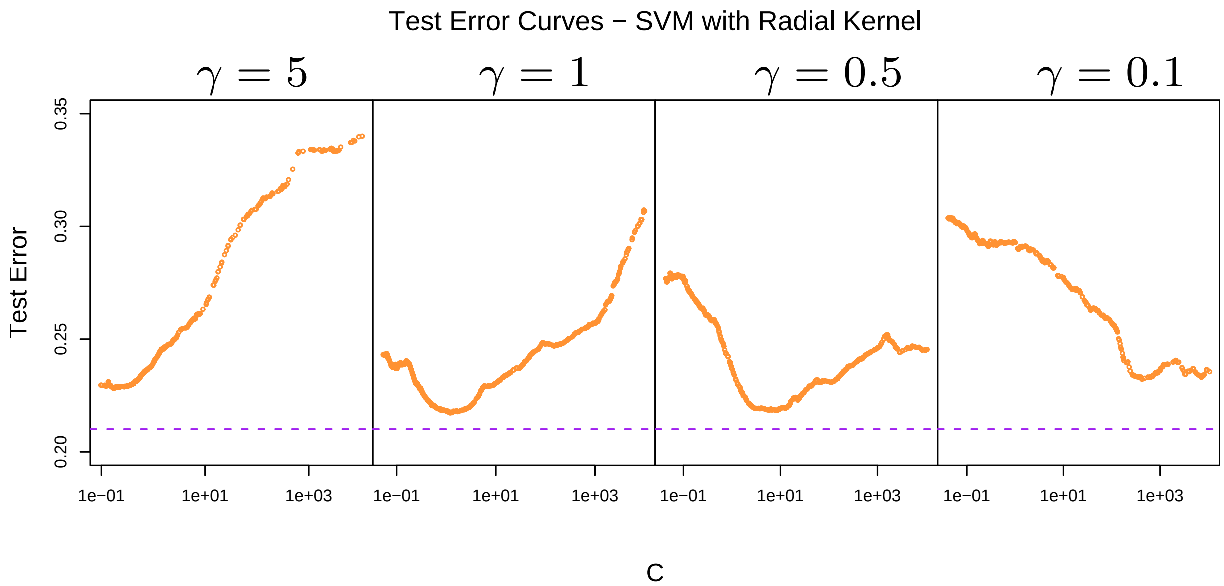



**FIGURE 12.6.** *Test-error curves as a function of the cost parameter $C$ for the radial-kernel SVM classifier on the mixture data. At the top of each plot is the scale parameter $\gamma$ for the radial kernel: $K_\gamma(x, y) = \exp -\gamma||x - y||^2$. The optimal value for $C$ depends quite strongly on the scale of the kernel. The Bayes error rate is indicated by the broken horizontal lines.*

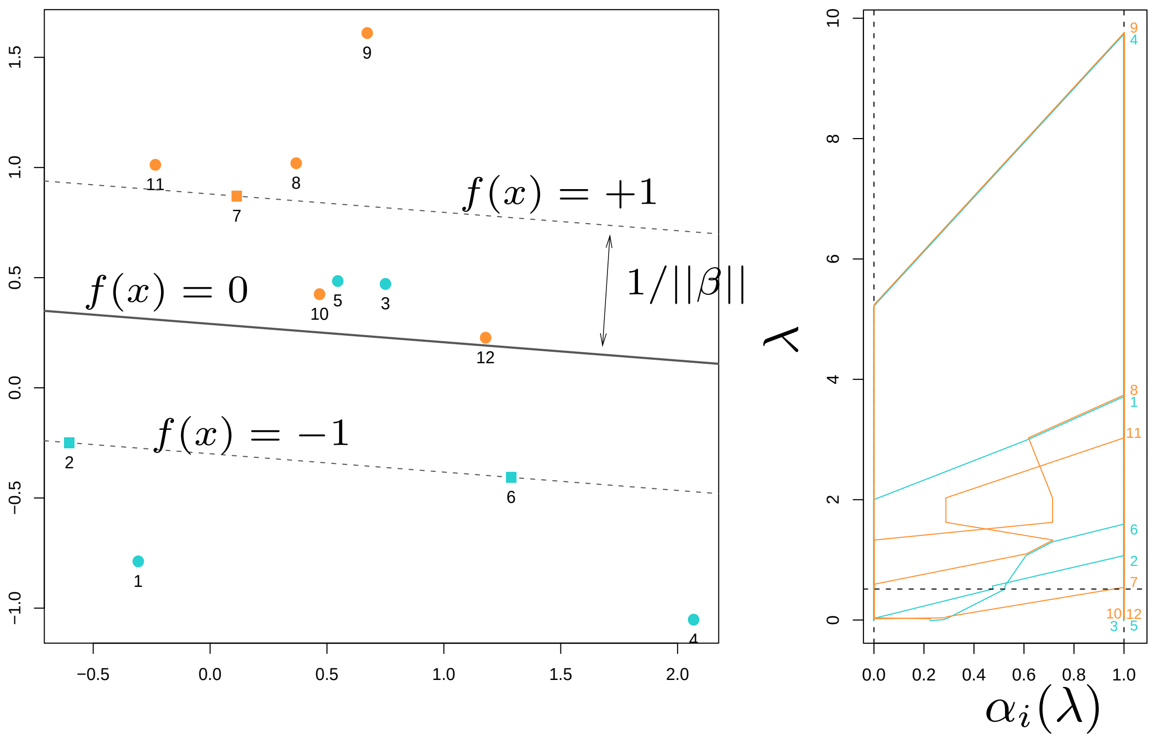



**FIGURE 12.7.** *A simple example illustrates the SVM path algorithm. (left panel:) This plot illustrates the state of the model at* $\lambda = 0.05$*. The "+1" points are orange, the "−1" blue.* $\lambda = 1/2$*, and the width of the soft margin is* $2/||\beta|| = 2 \times 0.587$*. Two blue points* $\{3, 5\}$ *are misclassified, while the two orange points* $\{10, 12\}$ *are correctly classified, but on the wrong side of their margin* $f(x) = +1$*; each of these has* $y_i f(x_i) < 1$*. The three square shaped points* $\{2, 6, 7\}$ *are exactly on their margins. (right panel:) This plot shows the piecewise linear profiles* $\alpha_i(\lambda)$*. The horizontal broken line at* $\lambda = 1/2$ *indicates the state of the* $\alpha_i$ *for the model in the left plot.*

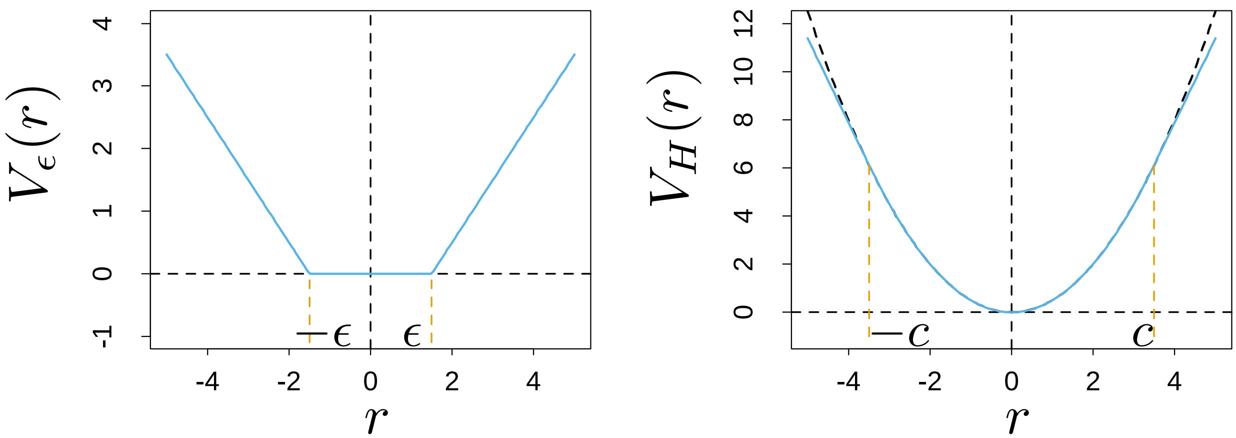



**FIGURE 12.8.** *The left panel shows the $\epsilon$-insensitive error function used by the support vector regression machine. The right panel shows the error function used in Huber's robust regression (blue curve). Beyond $|c|$, the function changes from quadratic to linear.*

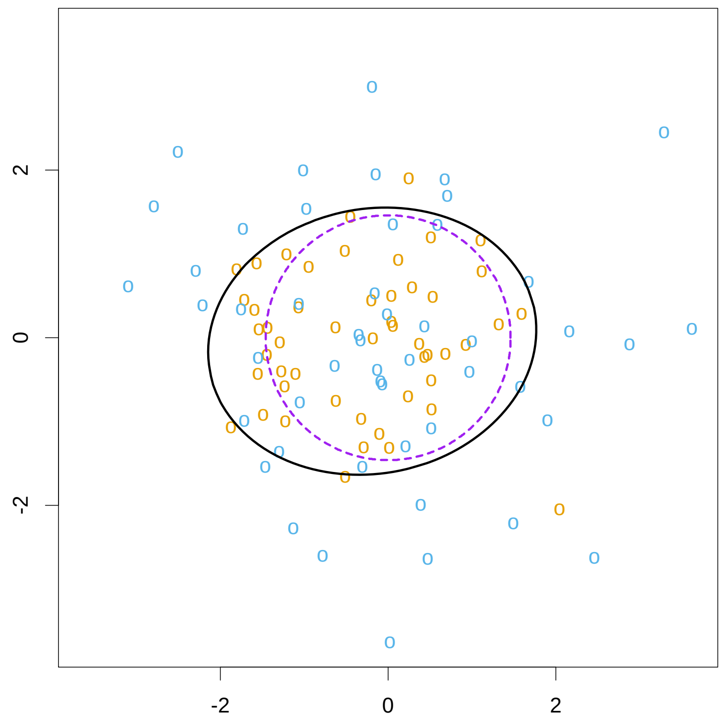



**FIGURE 12.9.** *The data consist of* 50 *points generated from each of* $N(0, I)$ *and* $N(0, \frac{9}{4}I)$. *The solid black ellipse is the decision boundary found by FDA using degree-two polynomial regression. The dashed purple circle is the Bayes decision boundary.*

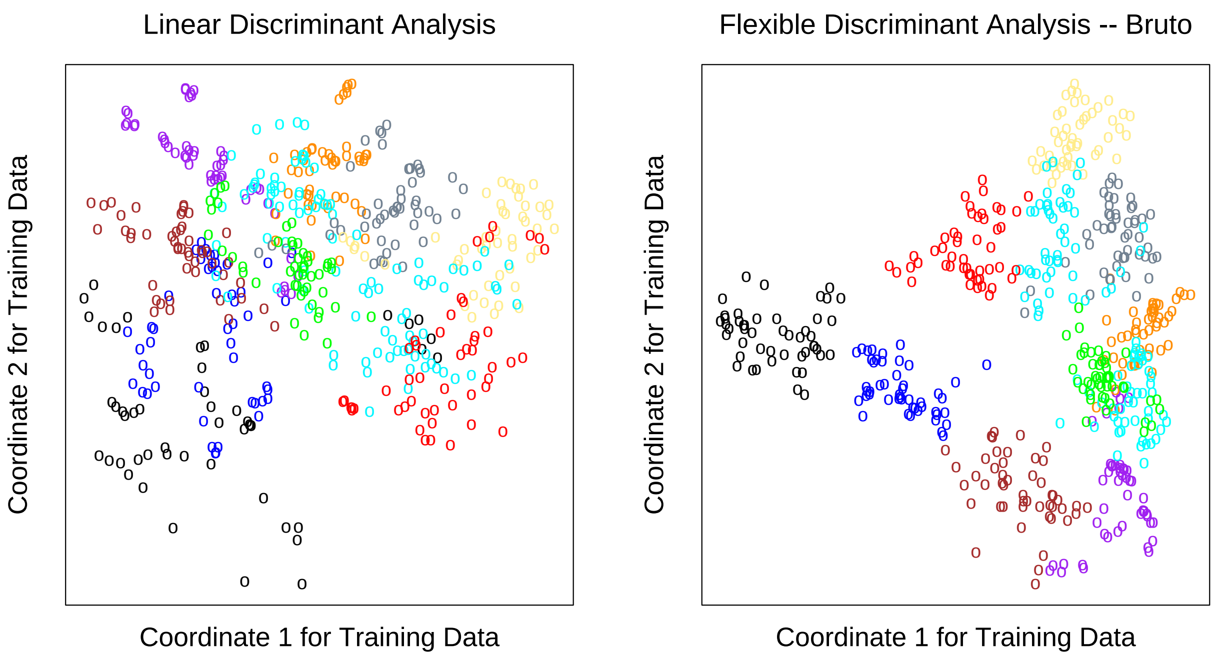



**FIGURE 12.10.** *The left plot shows the first two LDA canonical variates for the vowel training data. The right plot shows the corresponding projection when FDA/BRUTO is used to fit the model; plotted are the fitted regression functions $\hat{\eta}_1(x_i)$ and $\hat{\eta}_2(x_i)$. Notice the improved separation. The colors represent the eleven different vowel sounds.*

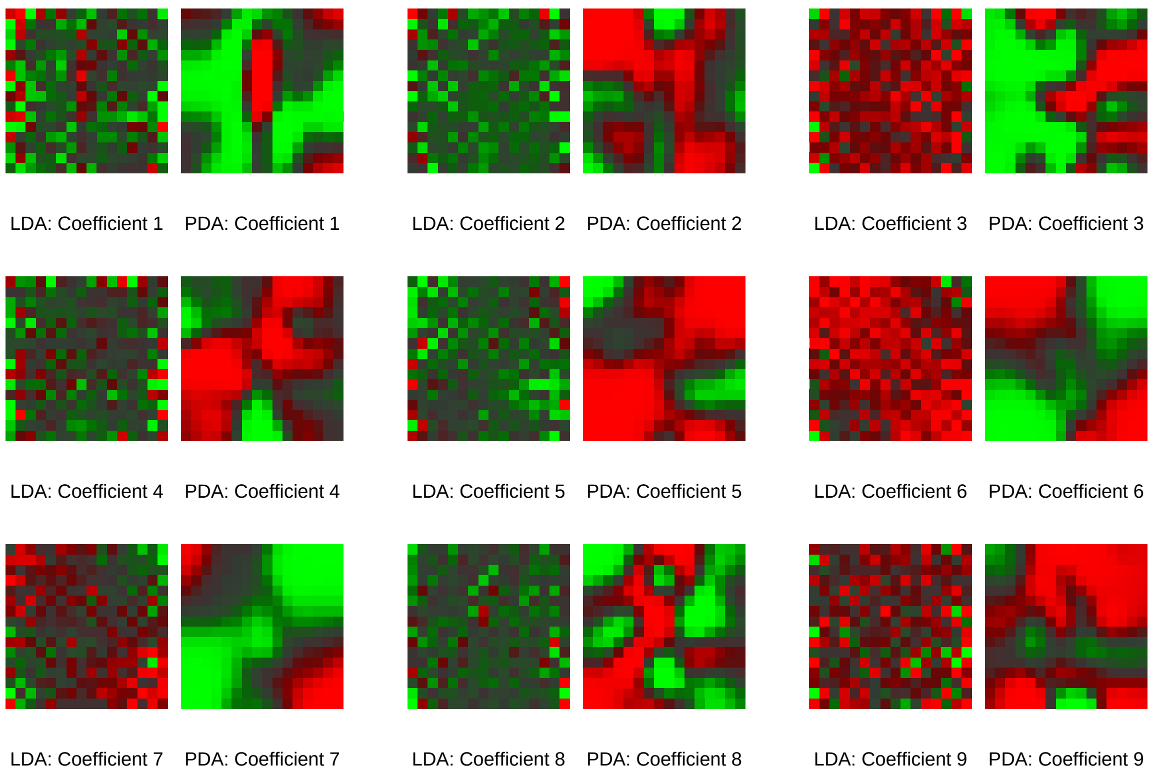



**FIGURE 12.11.** *The images appear in pairs, and represent the nine discriminant coefficient functions for the digit recognition problem. The left member of each pair is the LDA coefficient, while the right member is the PDA coefficient, regularized to enforce spatial smoothness.*

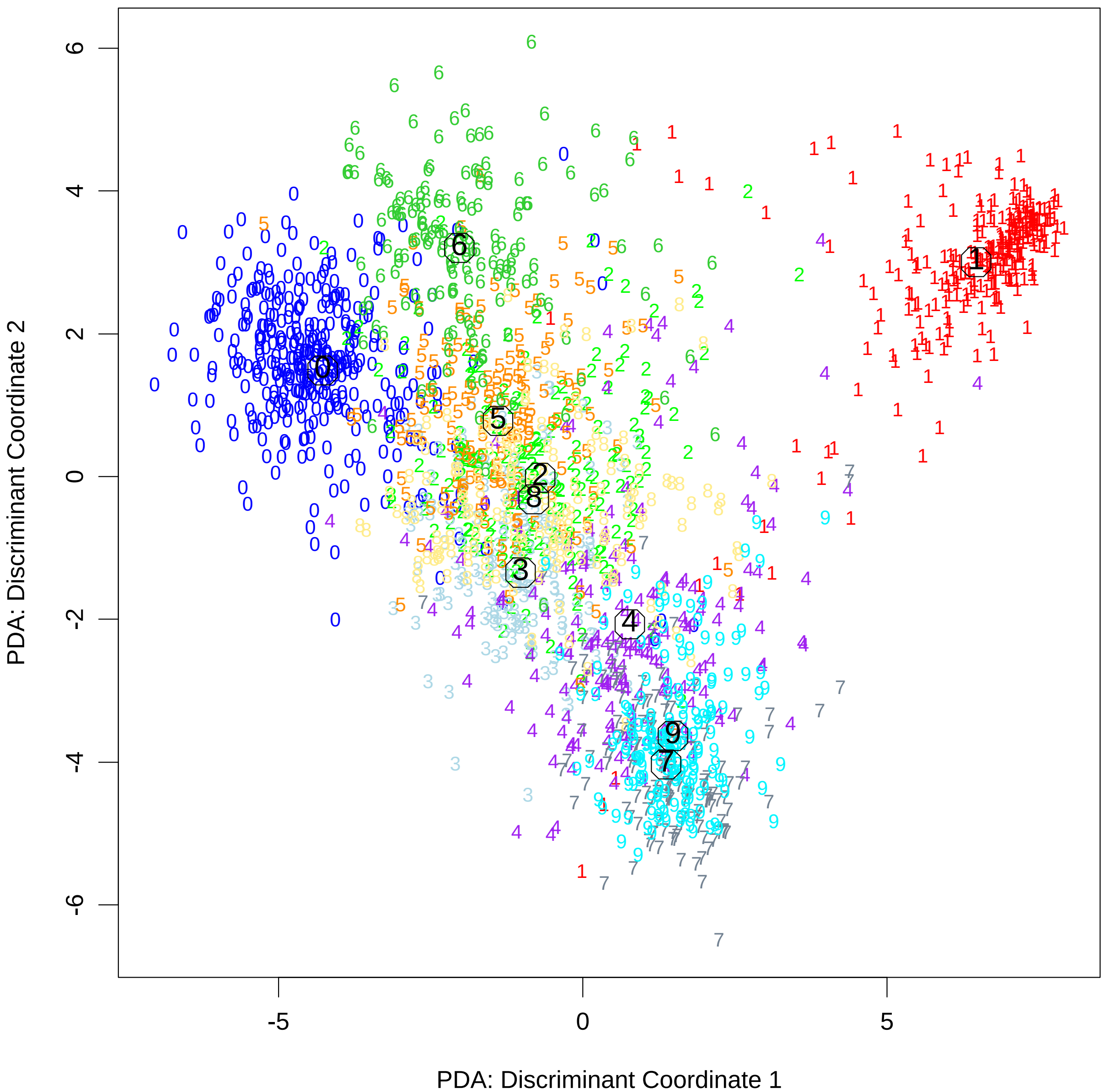
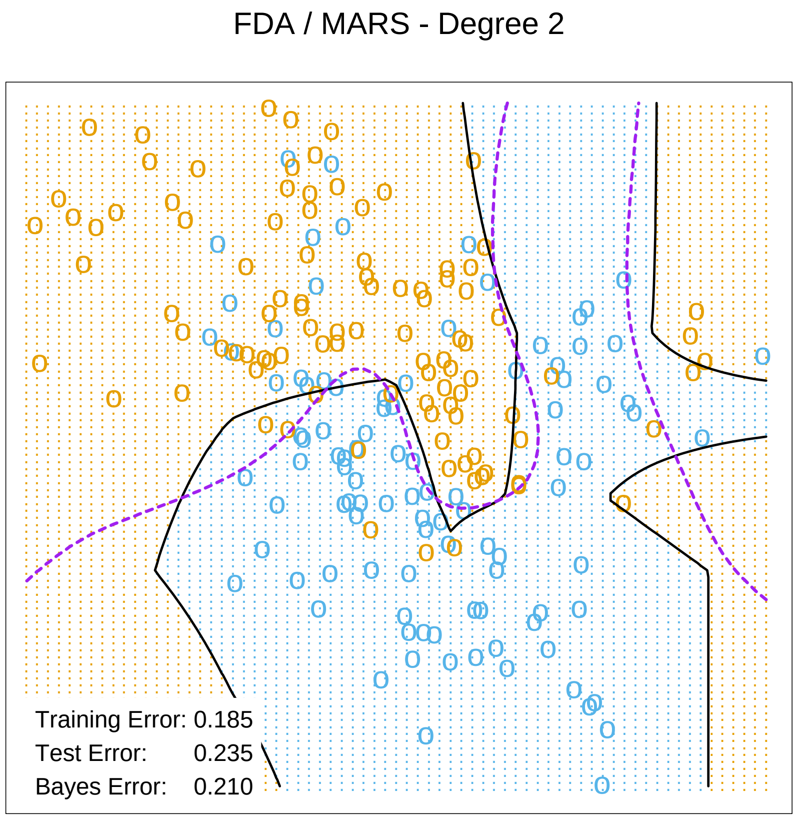



**FIGURE 12.12.** *The first two penalized canonical variates, evaluated for the test data. The circles indicate the class centroids. The first coordinate contrasts mainly 0's and 1's, while the second contrasts 6's and 7/9's.*



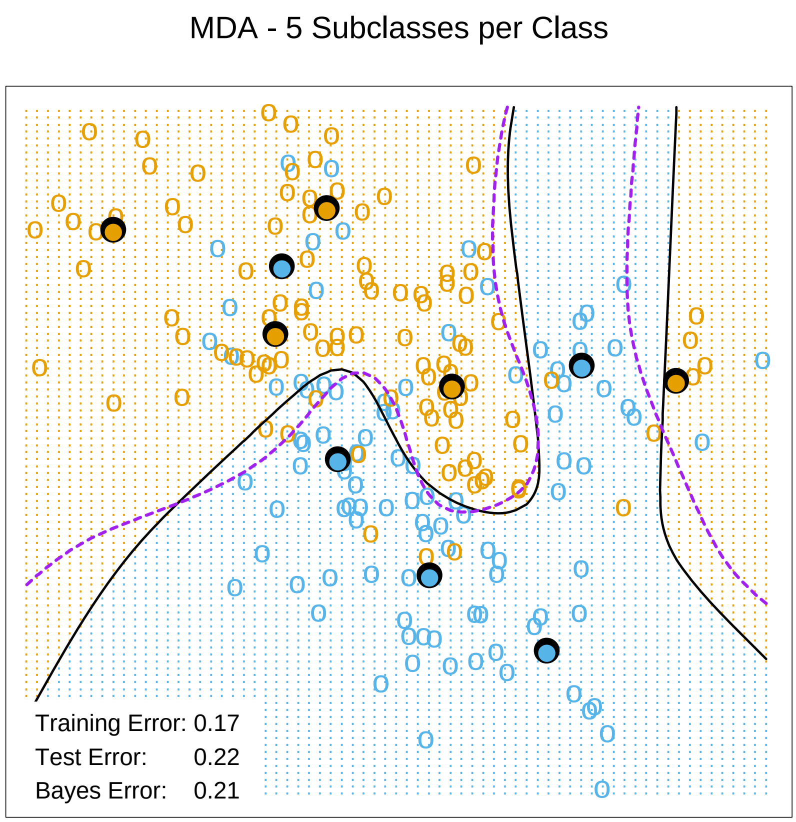




**FIGURE 12.13.** *FDA and MDA on the mixture data. The upper plot uses FDA with MARS as the regression*

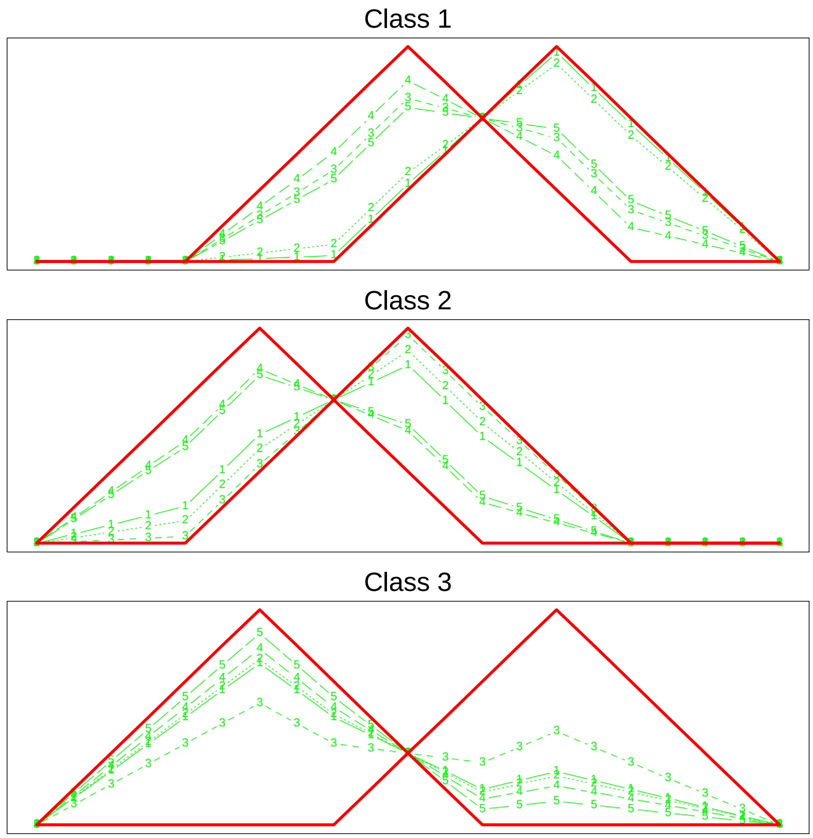



**FIGURE 12.14.** *Some examples of the waveforms generated from model (12.64) before the Gaussian noise is added.*

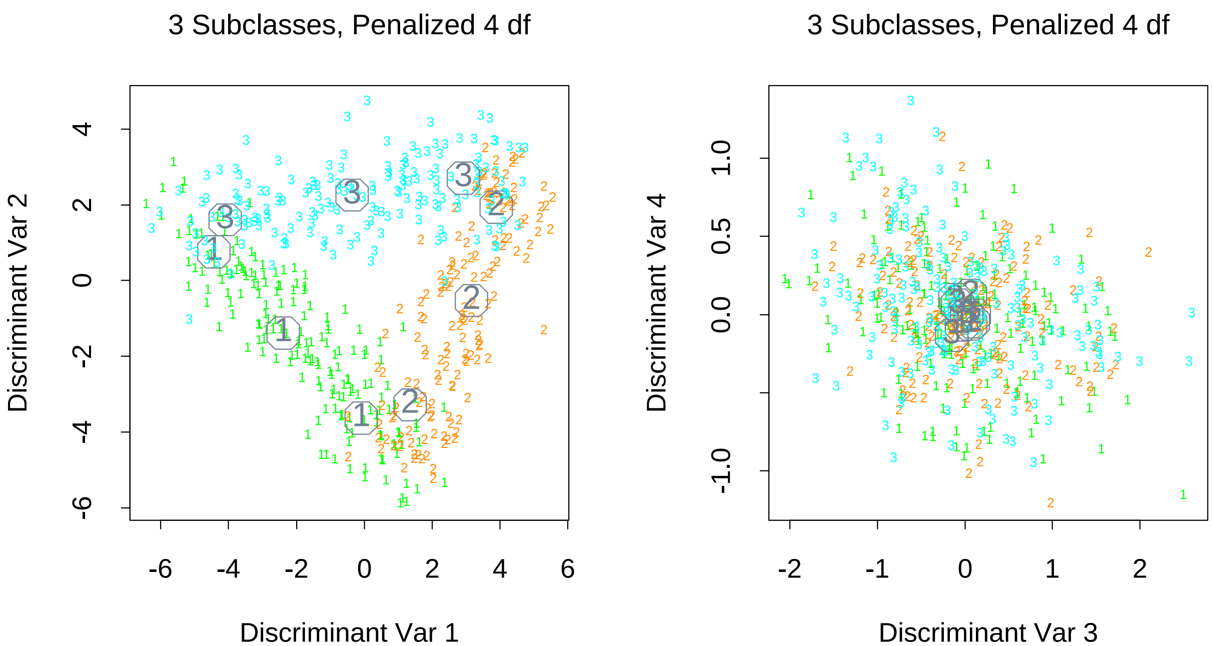



**FIGURE 12.15.** *Some two-dimensional views of the MDA model fitted to a sample of the waveform model. The points are independent test data, projected on to the leading two canonical coordinates (left panel), and the third and fourth (right panel). The subclass centers are indicated.*