rmix()

rmix() can be used to generate random samples from for mixture classification problems. It provides a non-linear example to check the performance of classifiers on.

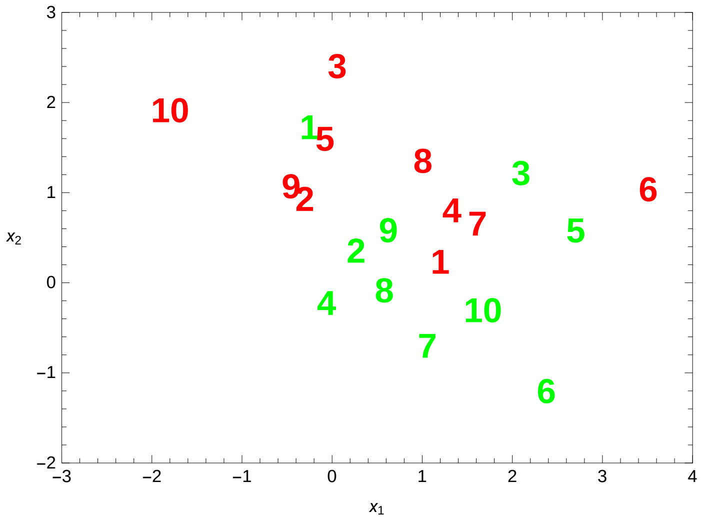
For the purpose of learning and understanding a new statistical method, simulation can be very helpful. We know what the correct result should be and we can generate an unlimited amount of data. After we learn about the performance of our method on some simulated examples, we can then experiment with real applications.

The green/red classes are generated by a mixture of bivariate normal distributions. For green and red the density functions can be written,

$\sum_{i=1}^{10} N(\mu_{G_i},\ \mathbb{I}_2/5)/10$

$\sum_{i=1}^{10} N(\mu_{R_i},\ \mathbb{I}_2/5)/10$

where $(\mu_{G_i},\ \mu_{R_i})$, $i = 1, \ldots, 10$ are shown in the plot below,



Within each class, each of the 10 normal distributions is chosen with equal probability. Again 100 from each class are used in the training sample.

i	$x_{i,1}$	$x_{i,2}$	y_i
1	2.14748	1.03185	0
2	3.36832	-0.181137	0
3	-0.247176	2.88748	0
4	1.00591	0.0900456	0
5	1.0272	-0.505474	0
...	...	...	...
196	-0.304882	0.944225	1
197	-0.230113	1.50992	1
198	-0.519905	0.94652	1
199	0.761488	0.388264	1
200	1.07735	1.82401	1

Here the classes are {0, 1} with $Y = 0/1$ corresponding to green/red respectively. Let $\mathrm{g}(x_1,\ x_2)$ and $\mathrm{r}(x_1,\ x_2)$ denote the pdf's for the green and red points. So

$\mathrm{g}(x_1,\ x_2) = \frac{1}{10} \sum_{i=1}^{10} \phi\left(x_1;\ g_{1,i},\ \frac{1}{5}\right) \phi\left(x_2;\ g_{2,i},\ \frac{1}{5}\right)$

where $\phi(x;\ \mu,\ \sigma^2)$ denotes the normal pdf with mean μ and variance σ^2 and $\mu_{G_i} = (g_{1,i},\ \ g_{2,i})$. Note that $\mathfrak{g}(x_1, x_2)$ and $\mathrm{r}(x_1, x_2)$ are the conditional probability densities corresponding to $\pi_q(x)$, $q = 0, 1$. Similarly for $\mathrm{r}(x_1, x_2)$. So the joint distribution may be written,

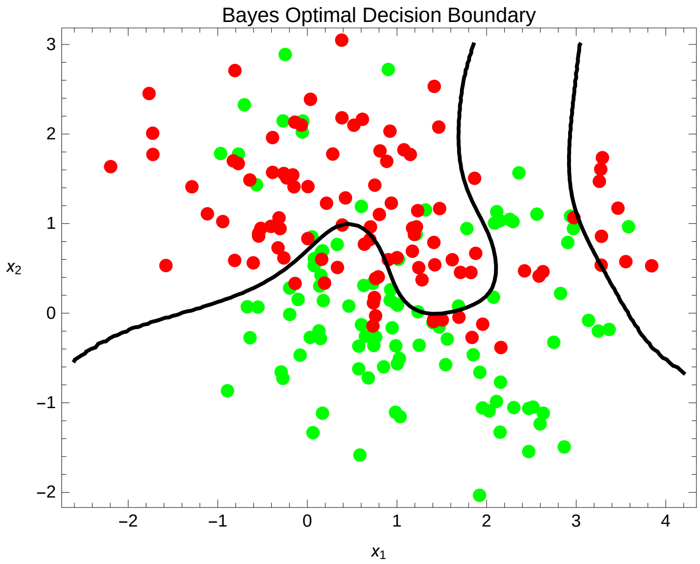
$\pi(x) = \mathfrak{g}(x_1, x_2) \Pr(Y = 0) + \mathrm{r}(x_1, x_2) \Pr(Y = 1)$

where $Y = 0$ represents green and $Y = 1$, red.

Assuming the prior probabilities for green and red are equal, so $\pi_0 = \pi_1 = 1/2$, the posterior probabilities are simply the joint densities divided by 2. So the Bayes classifier may be written,

$$\hat{y}(x_1, x_2) = \begin{cases} Y = 0 & \text{(green)} & \text{if } \mathfrak{g}(x_1, x_2) \ge \mathrm{r}(x_1, x_2) \\ Y = 1 & \text{(red)} & \text{if } \mathfrak{g}(x_1, x_2) < \mathrm{r}(x_1, x_2) \end{cases}$$

The boundary is determined by the contours of $\mathfrak{g}(x_1, x_2) = \mathrm{r}(x_1, x_2)$. This plot is constructed using ContourPlot[] in *Mathematica*.



Observed Mis-classification Rate = 21% on training data using the optimal Bayes decision boundary given in the above plot. On the training data, the classifier appears to have slightly weaker performance in the red region as shown in the confusion matrix.

	$\hat{Y}=0$	$\hat{Y}=1$
$Y=0$	83	17
$Y=1$	25	75

The Bayes error rate is the average or expected misclassfication rate. This could be estimated by simulation. In some cases we can compute this analytically.

$$\eta = \mathbb{E}\{\mathrm{EPE}\} = \mathbb{E}\left\{\mathfrak{L}\left(Y, \hat{\mathcal{Y}}(X_1, X_2)\right)\right\} \tag{1}$$

$$\eta = \int_{\mathbb{R}^2} (\Pr\{Y = 0 \mid X = x\} \times \mathfrak{L}(1, x) + \Pr\{Y = 1 \mid X = x\} \times \mathfrak{L}(0, x))\, \pi(x)\, dx \tag{2}$$

The joint probability density funciton of X is given by

$$\pi(x) = \pi(x_1, x_2) = \mathfrak{g}(x_1, x_2 \mid Y = 0) \Pr(Y = 0) + \mathrm{r}(x_1, x_2 \mid Y = 1) \Pr(Y = 1) \tag{3}$$

Assuming prior probabilties, $\pi(Y = 0) = \pi(Y = 1) = 1/2$,

$$\pi(x_1, x_2) = \frac{1}{2}(\mathfrak{g}(x_1, x_2 \mid Y = 0) + \mathrm{r}(x_1, x_2 \mid Y = 1)) \tag{4}$$

Also we have,

$$\Pr(Y=0 \mid X=x) = \frac{g(x)\Pr(Y=0)}{g(x)\Pr(Y=0)+r(x)\Pr(Y=1)} = \frac{g(x)}{g(x)+r(x)} \tag{5}$$

and

$$\Pr(Y=1 \mid X=x) = 1-\Pr(Y=0 \mid X=x).$$

The probability that a point is mis-classified is

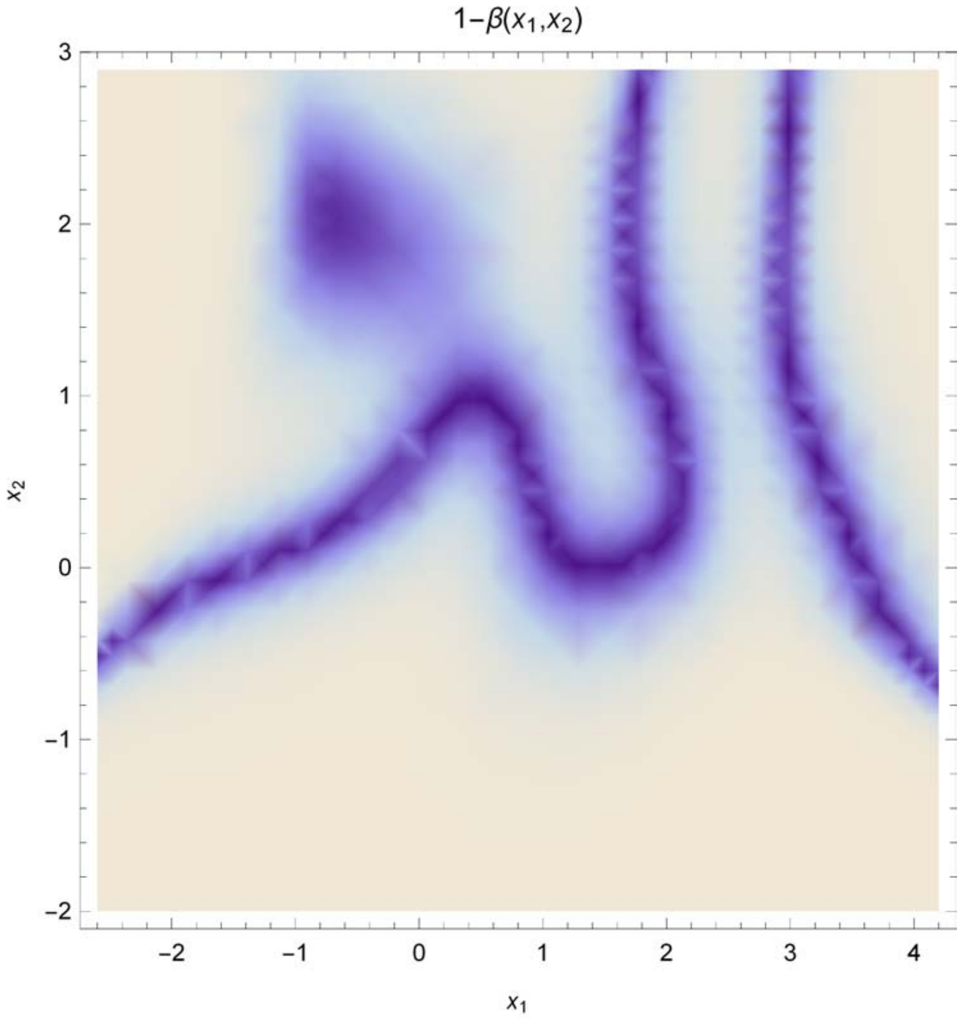
$$\beta(x) = \Pr(Y=0 \mid x)\,\mathcal{L}\big(1,\ \hat{Y}(x)\big) + \Pr(Y=1 \mid x)\,\mathcal{L}\big(0,\ \hat{Y}(x)\big) \tag{6}$$

Evaluating using numerical quadrature,

$$\eta = \int_{\mathbb{R}^2} \beta(x)\,\pi(x)\,dx = 20.76\,\% \tag{7}$$

This is the optimal misclassification rate for the best possible classifer and is called the Bayes rate. By comparison on test data with k-NN with $N = 2000$, the estimated mis-classification rate was 22.7% with a 95% C.I. (0.2086, 0.2454), so the true parameter is inside this interval.

The plot shows the probability of correct classification $1-\beta(x_1, x_2)$, where $\beta(x)$ is given in eqn. (6). We see that on the boundary and in the upper area between $-1.5 < x_1 < 0$ and $1.5 < x_2 < 2.5$, these probabilities are lowest.



Example

```
w
```