Proof that the Bayes Decision Rule is Optimal

**Theorem** For any decision function $g : \mathbb{R}^d \rightarrow \{0, 1\}$, 

$$\Pr\{g(X)! = Y\} \geq \Pr\{g^*(X)! = Y\}$$

We’ll prove it in the 2-class problem.

**Proof**

First we concentrate the attention on the error rate (probability of classification error) of the generic decision function $g(\cdot)$. Look at a SPECIFIC feature vector (namely, condition on $X = x$), and recall that uppercase letters denote a random variable, while a lowercase letter denotes a value.

$$\Pr\{g(X) \neq Y \mid X = x\} = 1 - \Pr\{g(X) = Y \mid X = x\}$$

so, when $X = x$ the probability of error is 1 minus the probability of correct decision. We make a correct decision if $g(X) = 1$ and $Y = 1$ OR if $g(X) = 0$ and $Y = 0$. Note that the events are disjoint, so the probability of the union (OR) is the sum of the probabilities

$$1 - \Pr\{g(X) = Y \mid X = x\} = 1 - \Pr\{g(X) = 1, Y = 1 \mid X = x\} - \Pr\{g(X) = 0, Y = 0 \mid X = x\}.$$ 

We now show that **conditional on** $X = x$, **the events** $\{g(X) = k\}$ and $\{Y = k\}$ **are independent** (surprising, isn’t it?).

First, note that conditional on $X = x$, $g(X) = g(x)$, and that, therefore, $g(x)$ is just the value of $g$ evaluated at $x$. This is either 0 or 1.

Assume WLOG that $g(x) = 1$. Then $\Pr\{g(x) = 0, Y = 0 \mid X = x\}$ is equal to zero, because $g(x)$ is equal to 1. Note, therefore, that the event $\{g(X) = 1\}$ has probability 0, and is conditionally independent of the event $Y = 0$ given $X = x$. Therefore:

$$\Pr\{g(X) = 0, Y = 0 \mid X = x\} = \Pr\{g(X) = 0 \mid X = x\} \Pr\{Y = 0 \mid X = x\}.$$ 

Similarly, $\Pr\{g(x) = 1, Y = 1 \mid X = x\} = \Pr\{Y = 1 \mid X = x\}$ because, by assumption, $\Pr\{g(X = 1) \mid X = x\} = 1$:

BUT an event having probability 1 is independent of any other event (can you prove it ?), then

$$\Pr\{g(X) = 1, Y = 1 \mid X = x\} = \Pr\{g(X) = 1 \mid X = x\} \Pr\{Y = 1 \mid X = x\}$$

by definition of independence.

Thus, for each $x$ where $g(x) = 1$,

$$\Pr\{g(X = k, Y = k \mid X = x)\} = \Pr\{g(X = k \mid X = x)\} \Pr\{Y = k \mid X = x\},$$

for $k = 0, 1$, and independence for this case is proved.
The same argument applies for each \( x \) where \( g(x) = 0 \): thus we can always write

\[
\Pr \{g(X) = k, Y = k \mid X = x\} = \Pr \{g(X) = k \mid X = x\} \Pr \{Y = k \mid X = x\},
\]

for \( k = 0, 1 \), which concludes the independence proof.

Now note that \( \Pr \{g(X) = k \mid X = x\} = 1 \) if \( g(x) = k \), and \( = 0 \) if \( g(x) \neq k \). By using the notation \( 1_A \) to denote the the indicator of the set \( A \), we can write:

\[
1 - \Pr \{g(X) = Y \mid X = x\} = 1 - (1_{g(x)=1} \Pr \{Y = 1 \mid X = x\} + 1_{g(x)=0} \Pr \{Y = 0 \mid X = x\}),
\]

Let’s now subtract \( \Pr \{g(X) = Y \mid X = x\} \) from \( \Pr \{g^*(X) = Y \mid X = x\} \):

\[
\Pr \{g^*(X) = Y \mid X = x\} - \Pr \{g(X) = Y \mid X = x\} = \Pr \{Y = 1 \mid X = x\} (1_{g^*(x)=1} - 1_{g(x)=1}) + \Pr \{Y = 0 \mid X = x\} (1_{g^*(x)=0} - 1_{g(x)=0})
\]

(simple algebra). Noting that \( \Pr \{Y = 0 \mid X = x\} = 1 - \Pr \{Y = 1 \mid X = x\} \), we can then write

\[
\Pr \{g^*(X) = Y \mid X = x\} - \Pr \{g(X) = Y \mid X = x\} = \Pr \{Y = 1 \mid X = x\} (1_{g^*(x)=1} - 1_{g(x)=1}) + (1 - \Pr \{Y = 1 \mid X = x\}) (1_{g^*(x)=0} - 1_{g(x)=0})
\]

(1)

Now, note that \( 1_{g^*(x)=0} = 1 - 1_{g^*(x)=1} \), etc. Hence,

\[
\Pr \{g^*(X) = Y \mid X = x\} - \Pr \{g(X) = Y \mid X = x\} = \Pr \{Y = 1 \mid X = x\} (1_{g^*(x)=1} - 1_{g(x)=1}) + (1 - \Pr \{Y = 1 \mid X = x\}) (1 - 1_{g^*(x)=1} - 1 + 1_{g(x)=1})
\]

\[
= (2\Pr \{Y = 1 \mid X = x\} - 1) (1_{g^*(x)=1} - 1_{g(x)=1})
\]

(2)

Now, note that, for each \( x \),

- if \( \Pr \{Y = 1 \mid X = x\} > 1/2 \), then by definition of the Bayes Decision Rule, \( 1_{g^*(x)=1} = 1 \), and, in general \( 1_{g(x)=1} \leq 1 \); thus, Eq 2 \( \geq 0 \).

- if \( \Pr \{Y = 1 \mid X = x\} < 1/2 \), then again by definition the Bayes Decision Rule, \( 1_{g^*(x)=1} = 0 \), and, in general \( 1_{g(x)=1} \geq 0 \); thus, Eq 2 \( \geq 0 \).

This is true for \( X = x \); Now, take the expectation with respect to \( f(X) \).