# Bayes optimal binary classification

- We consider the basic case of binary classification with 0-1 loss
- Then for a distribution $P(X, Y)$, the Bayes optimal classification $\hat{c}(x)$ for an instance $x \in \mathcal{X}$ becomes simply
  - if $P(Y = +1 \mid X = x) > P(Y = -1 \mid X = x)$ then $\hat{c}(x) = +1$
  - if $P(Y = +1 \mid X = x) < P(Y = -1 \mid X = x)$ then $\hat{c}(x) = -1$
  - if $P(Y = -1 \mid X = x) = P(Y = +1 \mid X = x)$ then we are free to choose either $\hat{c}(x) = +1$ or $\hat{c}(x) = -1$
- (this was derived on slide 95 from more general setting)

# Bayes optimal binary classification (2)

- Since $P(Y \mid X) = P(X, Y)/P(X)$ and $P(X)$ does not depend on $Y$, we can re-write this as
  - if $P(X = x, Y = +1) > P(X = x, Y = -1)$ then $\hat{c}(x) = +1$
  - if $P(X = x, Y = +1) < P(X = x, Y = -1)$ then $\hat{c}(x) = -1$
  - if $P(Y = -1, X = x) = P(Y = +1, X = x)$ then either $\hat{c}(x) = +1$ or $\hat{c}(x) = -1$ will do

- Finally, if we are given $P(Y)$ and $P(X \mid Y)$ instead of directly $P(X, Y)$, we can substitute
  - $P(X = x, Y = +1) = P(X = x \mid Y = +1)P(Y = +1)$
  - $P(X = x, Y = -1) = P(X = x \mid Y = -1)P(Y = -1)$

# Bayes error for binary classification

- Generally, for a binary classifier $\hat{c}$ we can write its expected 0-1 loss, i.e. probability of making a mistake, as

$$\sum_{x:\hat{c}(x)=-1} P(x,+1) + \sum_{x:\hat{c}(x)=+1} P(x,-1)$$

(where we now write simply $P(x,y)$ instead of $P(X=x, Y=y)$)

- If $\hat{c}$ is the Bayes optimal classifier, we have $\hat{c}=-1$ if $P(x,+1) < P(x,-1)$, and $\hat{c}=+1$ if $P(x,-1) < P(x,+1)$

- Therefore, we can write the Bayes error as

$$\sum_{x\in\mathcal{X}} \min\{\, P(x,+1), P(x,-1)\,\}$$

# Continuous case

- In the continuous-valued case $x \in \mathbb{R}^d$ (with $Y$ still binary) we have continuous conditional densities $p(x \mid Y = +1)$ and $p(x \mid Y = -1)$ which replace conditional probabilities $P(x \mid Y = +1)$ and $P(x \mid Y = -1)$
- Everything works as before:
  - $\hat{c}(x) = +1$ if $p(x \mid Y = +1)P(Y = +1) > p(x \mid Y = -1)P(Y = -1)$
  - Bayes error is

$$\int_{x \in \mathcal{X}} \min(p(x, +1), p(x, -1))\, dx$$

- The *decision boundary* is the set

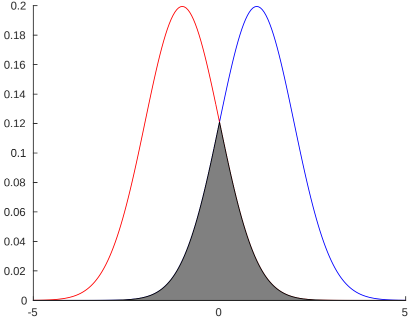
$$\{\, x \in \mathcal{X} \mid p(x, +1) = p(x, -1) \,\}$$

# Bayesian classifier in one dimension

- In one-dimensional case $x \in \mathbb{R}$ this becomes particularly clear
  - plot curves $p(x \mid y = +1)P(Y = +1)$ and $p(x \mid y = -1)P(Y = -1)$
  - decision boundary is given by the interection point(s) of the two curves
  - Bayes error is the area remaining below the lower curve

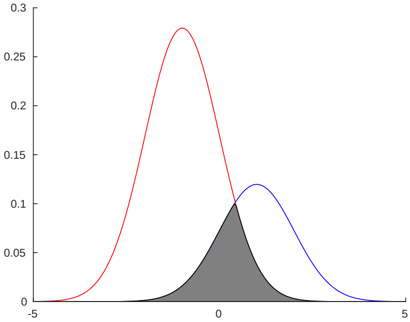
# Example 1: similar Gaussians

- uniform prior: $P(Y = +1) = P(Y = -1) = 1/2$
- Gaussian density $p(x \mid Y = y) = \mathcal{N}(\mu_y, \sigma_y^2)$ where $\mu_+ = 1$, $\mu_- = -1$ and $\sigma_+^2 = \sigma_-^2 = 1$
- easy to see that decision boundary is the middle point between the class means



# Example 2: Gaussians with non-uniform prior

- now $P(Y = +1) = 0.3$ and $P(Y = -1) = 0.7$
- again $p(x \mid Y = y) = \mathcal{N}(\mu_y, \sigma_y^2)$ where $\mu_+ = 1$, $\mu_- = -1$ and $\sigma_+^2 = \sigma_-^2 = 1$
- easy to see that decision boundary shifts towards the less prevalent class



# Example 3: Gaussians with different variance

- again $P(Y = +1) = P(Y = -1) = 1/2$
- again Gaussians $p(x \mid Y = y) = \mathcal{N}(\mu_y, \sigma_y^2)$
  - $\mu_+ = 1$ and $\mu_- = -1$
  - $\sigma_+^2 = 1$ and $\sigma_-^2 = 2$
- now the decision boundary consists of two points