Classification

- **Bayesian Classifier**

The Bayesian Classifier...Bayes Theorem

In probability theory, **Bayes' theorem** (often called **Bayes' Law**) relates the conditional and marginal probabilities of two random events. It is often used to compute posterior probabilities given observations.

For any events, A and B,

Marginal probabilities: P(A), P(B).

Conditional probabilities: P(A|B), P(B|A).

The Bayesian Classifier...Bayes Theorem

For A and B two events,

$$P(A|B) = \frac{P(A\&B)}{P(B)} \Rightarrow P(A|B)P(B) = P(A\&B)$$

$$P(B|A) = \frac{P(A\&B)}{P(A)} \Rightarrow P(B|A)P(A) = P(A\&B)$$

So,

$$P(A|B)P(B) = P(B|A)P(A)$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}$$

The above formula is referred to as *Bayes' theorem*. It is extremely useful in decision analysis.

The Bayesian Classifier...Bayes Theorem

Consider each attribute and class label as random variables

Given a record with attributes ($A_1, A_2,\ldots,A_n$)

- Goal is to predict class C (= ($c_1, c_2, \ldots, c_m$))
- Specifically, we want to find the value of C that maximizes
  - $P(C| A_1, A_2,\ldots,A_n)$

Can we estimate $P(C| A_1, A_2,\ldots,A_n)$ directly from data?

The Bayesian Classifier...Bayes Theorem

Approach:

compute the posterior probability $P(C \mid A_1, A_2, \ldots, A_n)$ for all values of C using the Bayes theorem

$$\boxed{P(C \mid A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n \mid C) P(C)}{P(A_1 A_2 \ldots A_n)}}$$

Choose value of C that maximizes

$$P(C \mid A_1, A_2, \ldots, A_n)$$

Equivalent to choosing value of C that maximizes

$$P(A_1, A_2, \ldots, A_n \mid C)\, P(C)$$

How to estimate $P(A_1, A_2, \ldots, A_n \mid C)$?

The conditional probabilities are difficult to evaluate if the number of variables are more.

The Bayesian Classifier... Bayes Theorem

Assume independence among features/attributes A_i when class is given:

$$P(A_1, A_2, \ldots, A_n | C) = P(A_1 | C_j)\, P(A_2 | C_j) \ldots P(A_n | C_j)$$

Can estimate $P(A_i | C_j)$ for all A_i and C_j?

New point is classified to C_j if $P(C_j)\, \Pi\, P(A_i | C_j)$ is maximal.

Note: The above is equivalent to find i such that $\Pi\, P(A_i | C_j)$ is maximal, since $P(C_j)$ is identical.

Due to assumption of **independence** of features, we call Bayes classifies as **Naïve Bayes classifier**.

The Bayesian Classifier...Continuous features

For continuous attributes:

- Discretize the range into bins
  one ordinal attribute per bin
  violates independence assumption
- Two-way split: $(A < v)$ or $(A > v)$
  choose only one of the two splits as new attribute
- Probability density estimation:
  Assume attribute follows a normal distribution
  Use data to estimate parameters of distribution (e.g., mean and standard deviation)
  Once probability distribution is known, can use it to estimate the conditional probability $P(A_i|c)$

The Bayesian Classifier...

- It is possible to show that the test error rate of Bayesian classifier can be minimized.
- This very simple classifier.
- In a two-class problem where there are only two possible response values, say class 1 or class 2, the Bayes classifier.
- Predicts class one if $\Pr(Y = 1|X = x_0) > 0.5$, and class two otherwise.

The Bayesian Classifier...

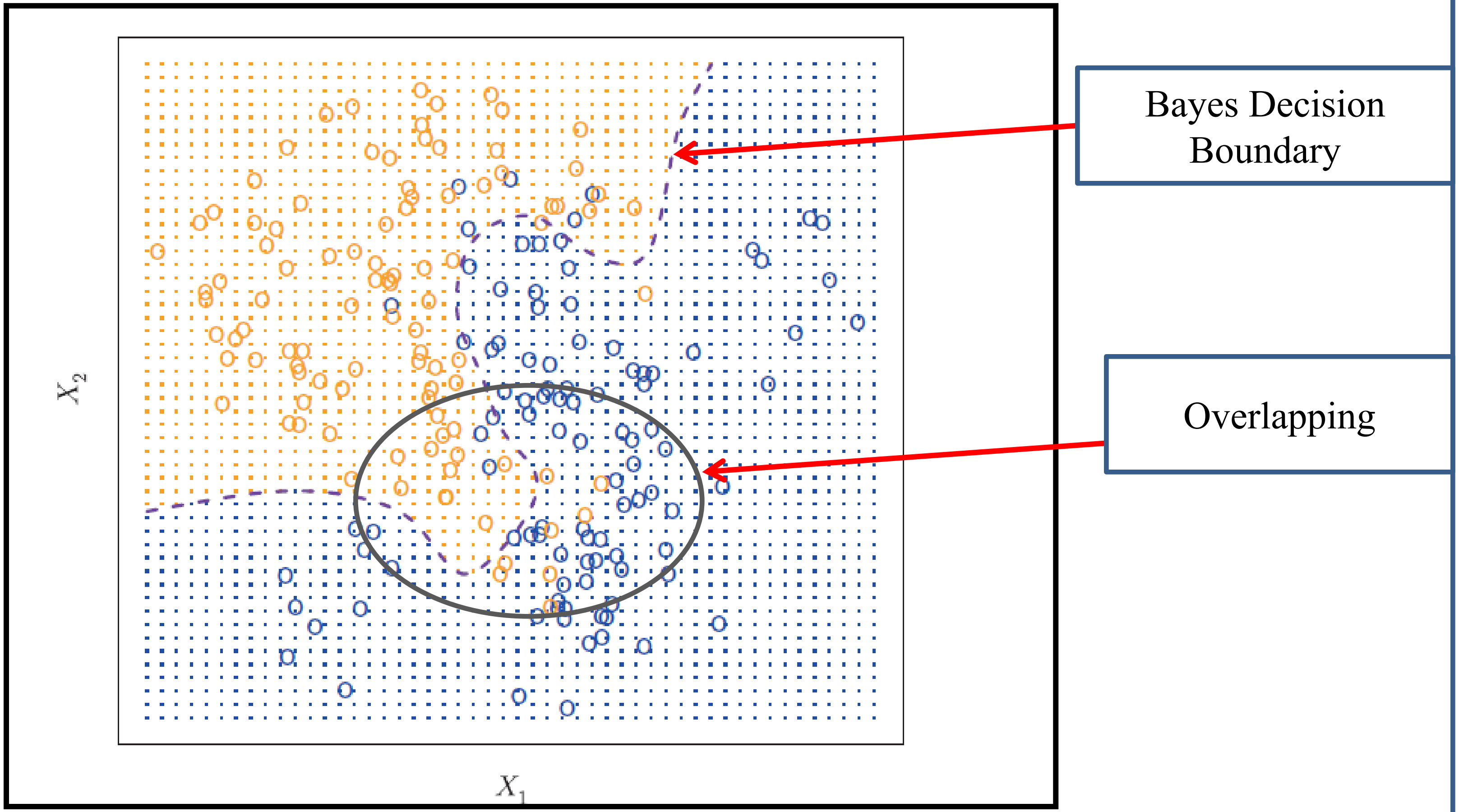
- The Bayes classifier produces the lowest possible test error rate, called the Bayes error rate. Since the Bayes classifier will always choose the class for which the conditional probability is largest.
- The error rate at $X = x_0$ will be

$$1 - \underset{j}{\mathrm{Max}}\, \Pr\{Y = j \mid X = x_0\}$$

- In general, the overall Bayes error rate is given by

$$1 - E\left(\underset{j}{\mathrm{Max}}\, \Pr\{Y = j \mid X = x_0\}\right)$$

The Bayesian Classifier...



In the above fig. we have two classes indicated by 'orange' and 'blue' colors.

The Bayesian Classifier...

- The purple dashed line (Bayes decision boundary) represents the points where the probability is exactly 50%.
- The Bayes classifier's prediction is determined by the Bayes decision boundary.
- In above given case, the overlapping may cause misclassification and calculated as Bayes error rate.

Thank You