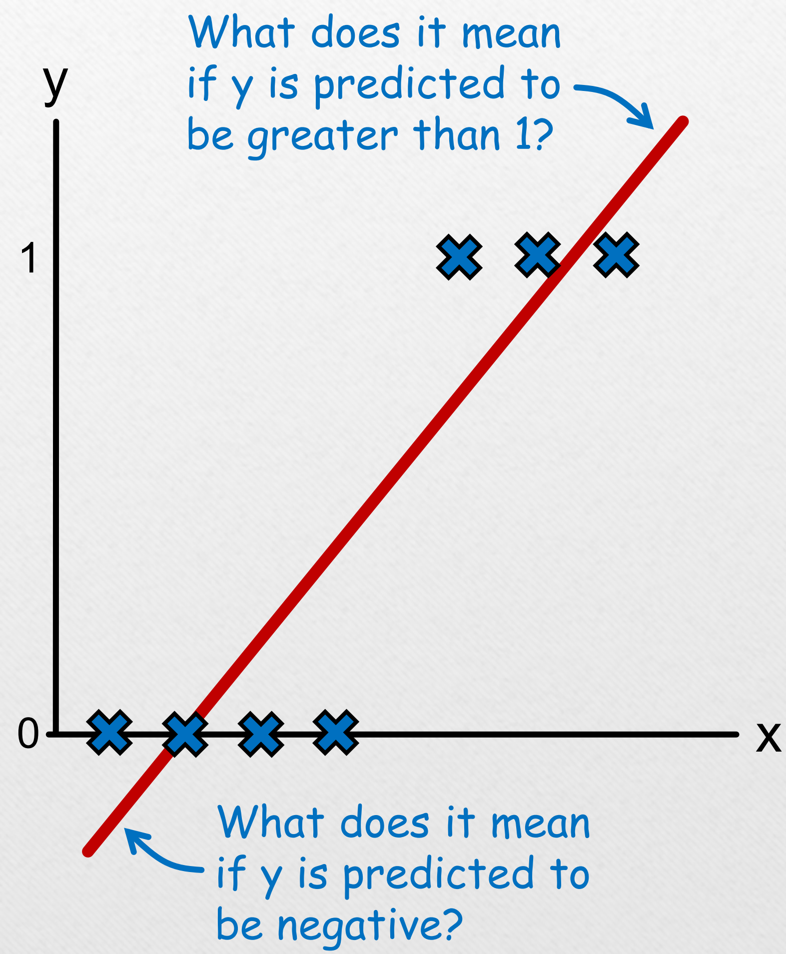
Classification

Classification

- Classification is the task of predicting a **discrete dependent variable**
- E.g. $y = 1$ if married, $y = 0$ otherwise
- E.g. $y = 1$ if studying in CUHK, $y = 2$ if studying in HKU, $y = 3$ if studying in HKUST
- Note that the term "classification" is only commonly used in machine learning

What's the Problem with OLS?

- The problem is that OLS's predictions can take on any real number, even though we know the actual values are discrete
- We want models that explicitly give discrete predictions
- Ideas?

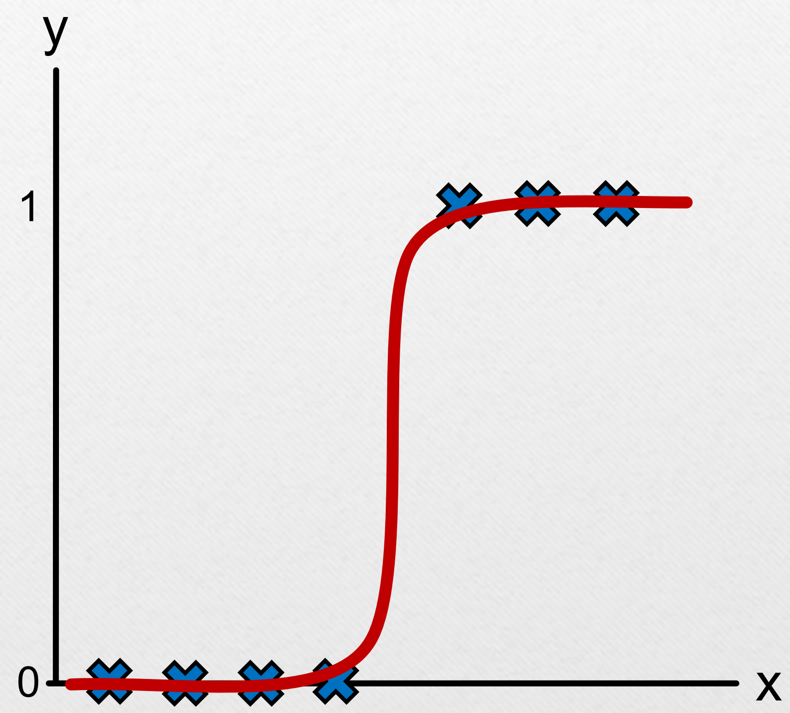


Logistic Regression

- A logistic regression—or **logit** for short—models a binary dependent variable in the following way:

$$\Pr(y = 1|X) = \frac{e^{X\vec{\beta}}}{1 + e^{X\vec{\beta}}}$$

- This function is called the **logistic function** or the **sigmoid function**

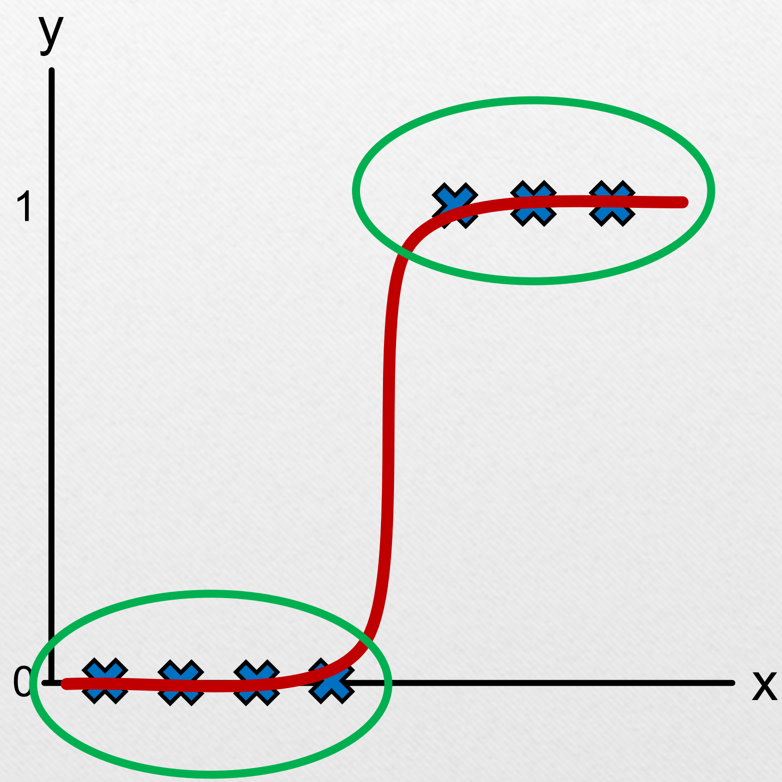


Logistic Regression

$$\Pr(y = 1|X) = \frac{e^{X\vec{\beta}}}{1 + e^{X\vec{\beta}}}$$

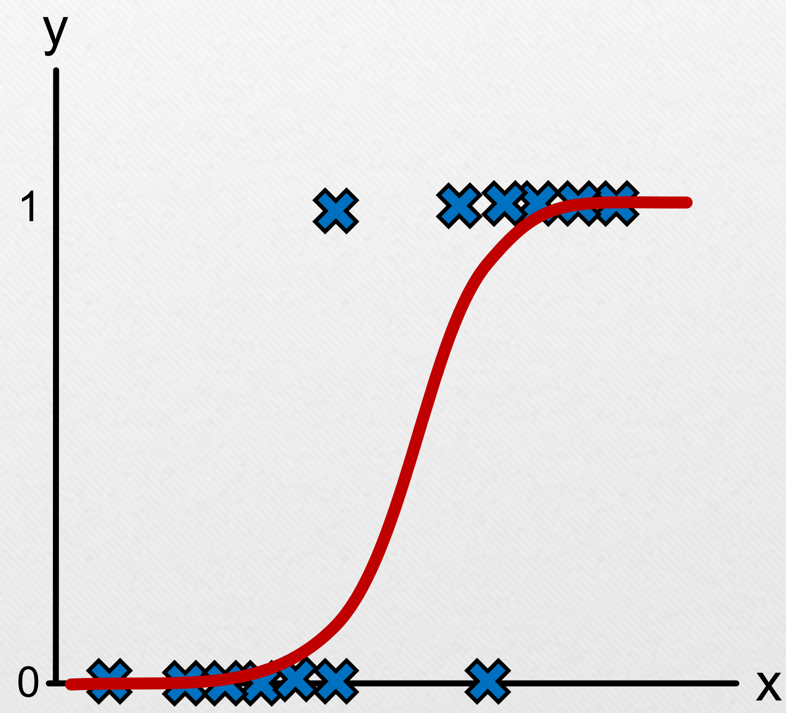
This function is called the **logistic function** or the **sigmoid function**

- If $X\vec{\beta}$ is large, $e^{X\vec{\beta}}$ is large, so $\frac{e^{X\vec{\beta}}}{1+e^{X\vec{\beta}}} \approx \frac{e^{X\vec{\beta}}}{e^{X\vec{\beta}}} = 1$
- If $X\vec{\beta}$ is small, $e^{X\vec{\beta}} \approx 0$, so $\frac{e^{X\vec{\beta}}}{1+e^{X\vec{\beta}}} \approx \frac{0}{1} = 0$



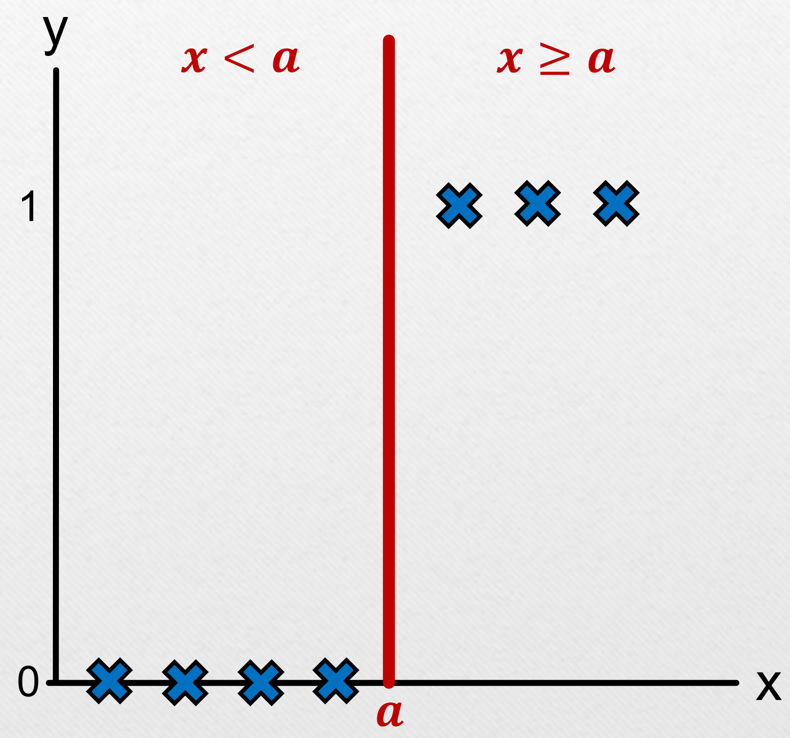
Logistic Regression

- A logistic regression can accommodate data without a clear-cut split by widening the transition region



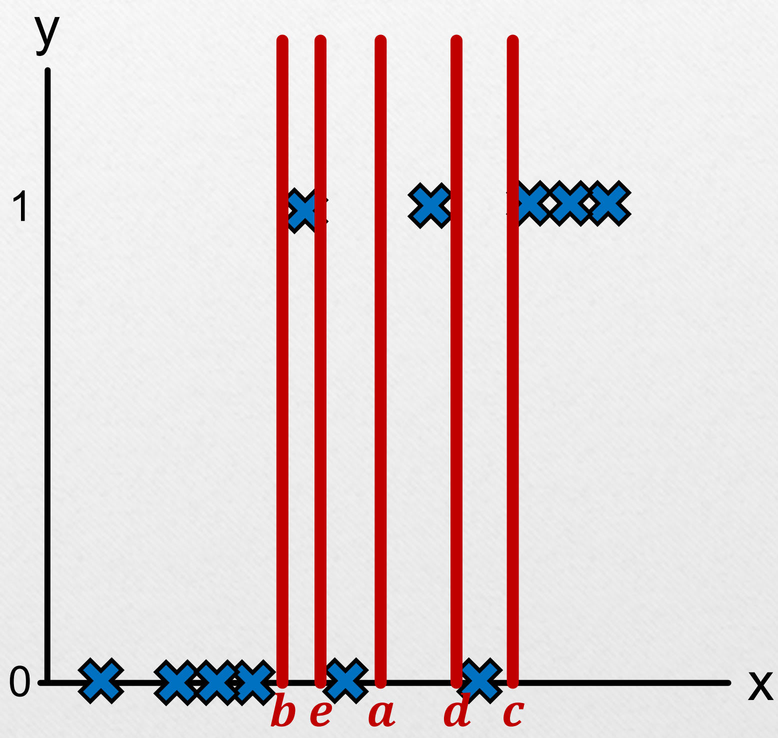
Decision Tree

- **Decision tree** explains the dependent variable by partitioning the independent variables based on inequality conditions



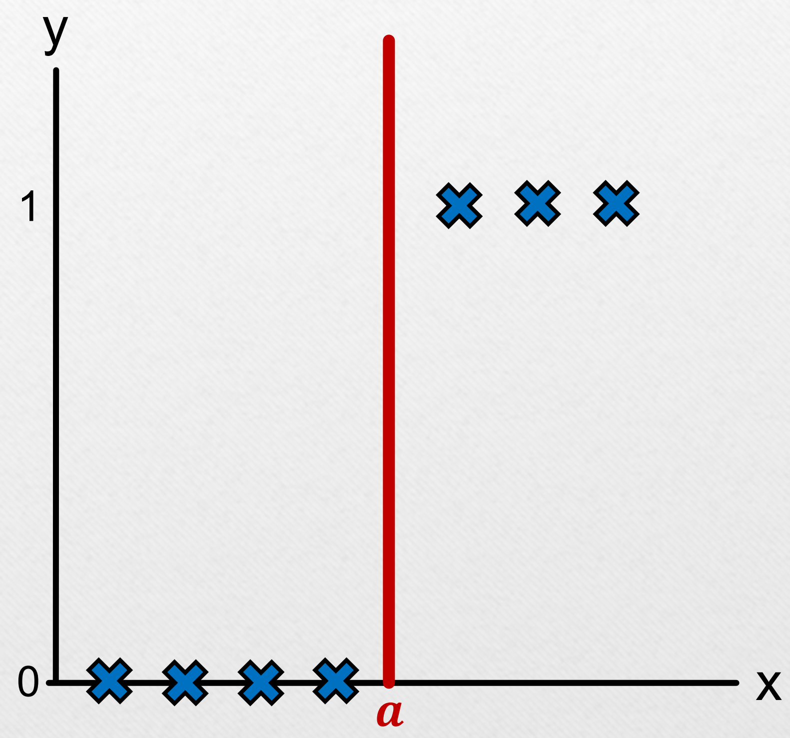
Decision Tree

- Decision tree can explain complex relationships by increasing the number of partitions
- Risk of overfit!



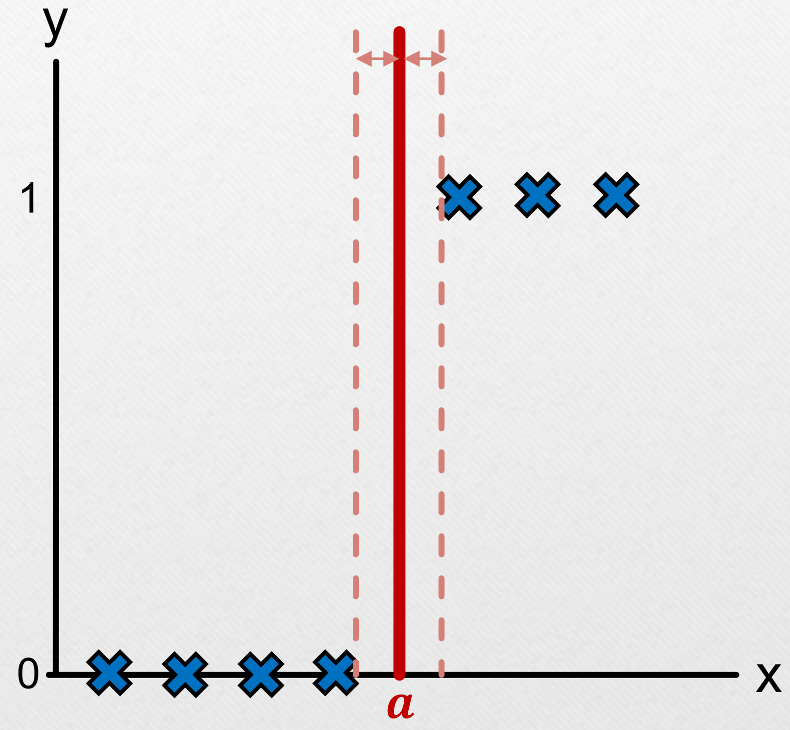
Support Vector Machine

- **Support vector machine** seeks to find a line that splits the dependent variable according to its value
- When there is more than one independent variable, what we have is a (multi-dimensional) plane instead of a line.



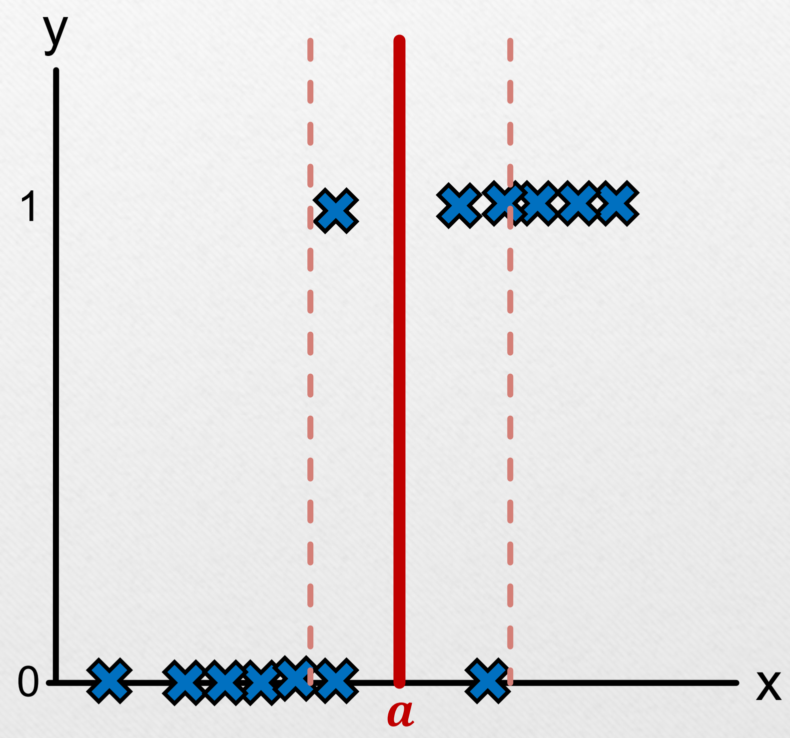
Support Vector Machine

- **Support vector machine** (SVM) seeks to find a line that splits the dependent variable according to its value
- Unlike regression tree, this line is not defined based on an inequality condition, but rather found by maximizing the distance between the line and the points on each side
- The dotted lines are called the **margins**



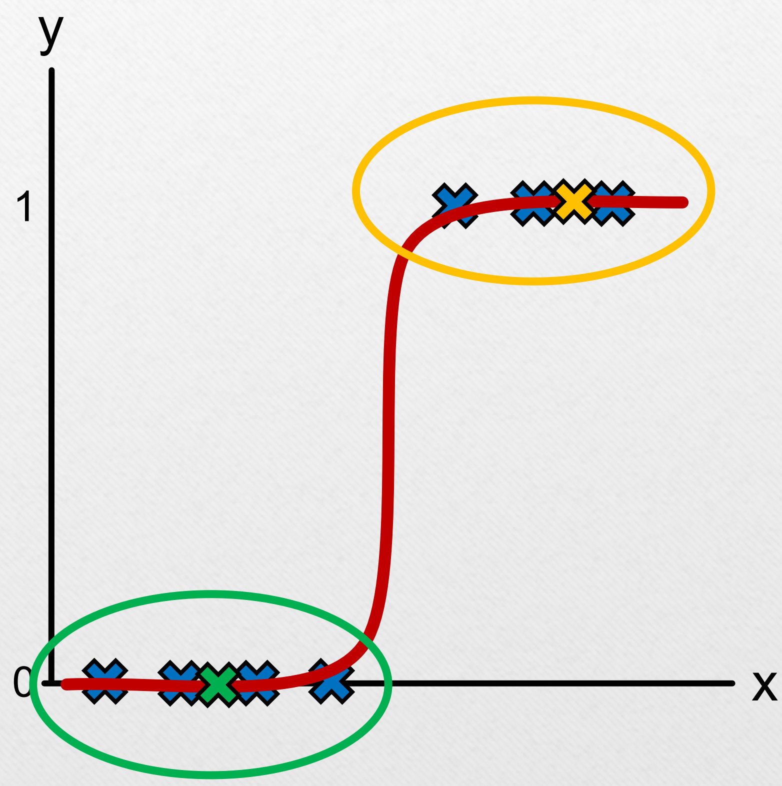
Support Vector Machine

- SVM can accommodate data without a clear-cut split by allowing for a soft margin



Nearest Neighbor

- Nearest neighbor classifies each observation with the same label as the observation in the training set that most resemble it
- To prevent overfitting, use the mode of several dozen neighbors instead of just the closest one.



Naïve Bayes

Bayes Rule: $$P(y|\vec{x}) = \frac{P(\vec{x}|y)P(y)}{P(\vec{x})}$$

$$P(y|\vec{x}) \propto P(\vec{x}|y)P(y)$$

$P(y)$ can be computed from data. Need $P(\vec{x}|y)$.

Naïve Bayes assumes

$$P(\vec{x}|y) = P(x_1|y) \cdot P(x_2|y) \cdot P(x_3|y) \cdot \ldots$$

Naïve Bayes is popularized by its use in some of the earliest email spam filters.