Supervised Machine Learning

Supervised Learning

Your "training" dataset is composed of examples of labeled examples:

- - features
- - labels

A supervised learning model learns to predict the label from the features .

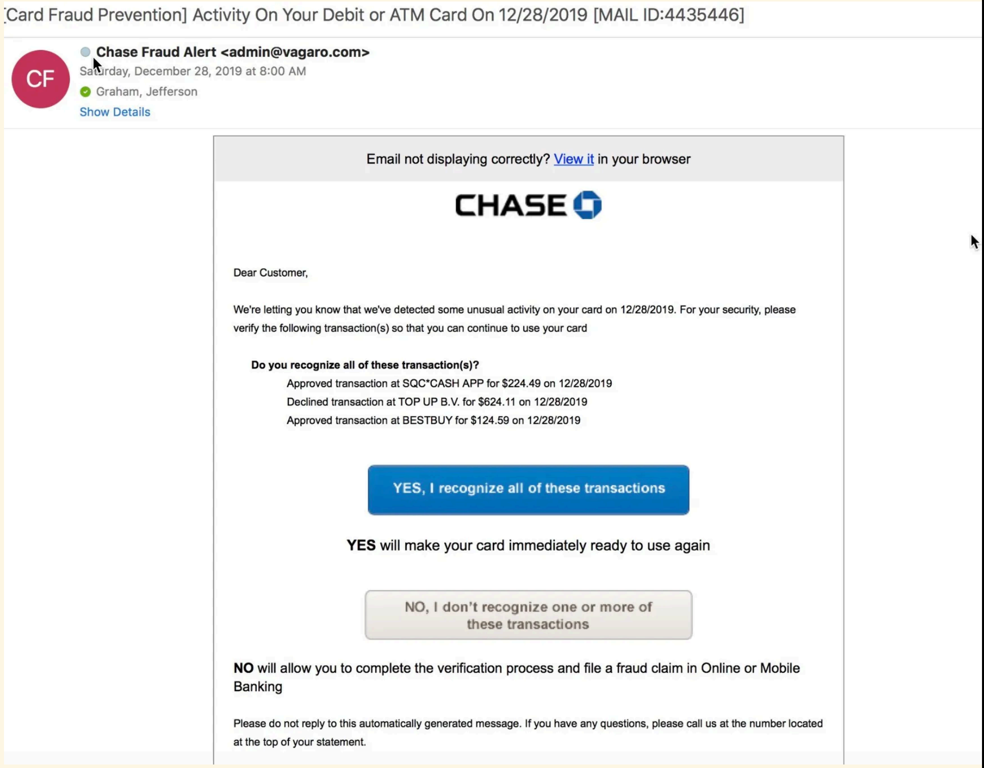
Supervised Learning Examples

Credit Card Fraud Detection

- **Features:** Vendor, location, time, distance from last transaction
- **Labels:** Chargebacks on previous transactions

Question: What kind of supervised learning?

1. Regression
2. Binary Classification
3. Multiclass Classification



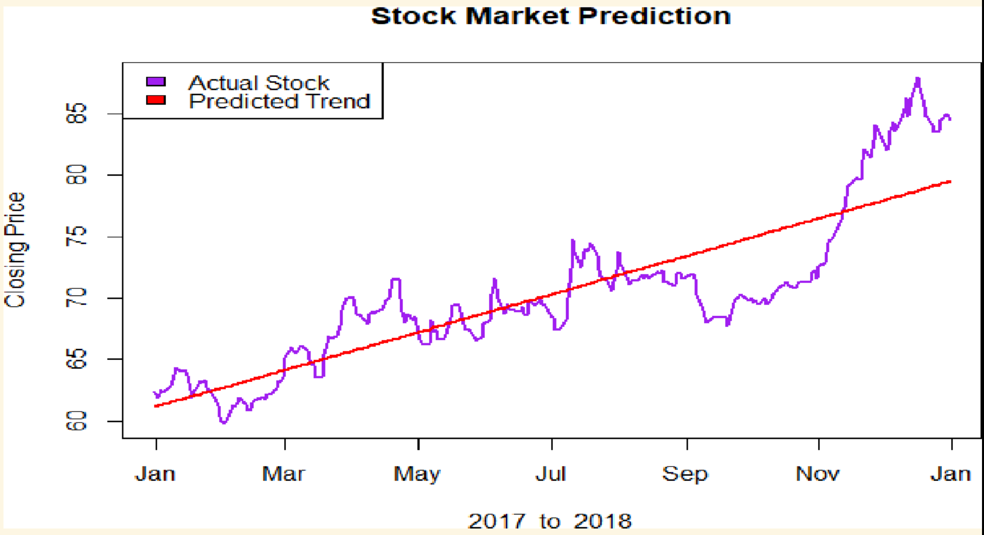
Supervised Learning Examples

Stock Market Prediction

- **Features:** Stock price from Feb 1st to March 1st
- **Labels:** Stock price on March 7th.

Question: What kind of supervised learning?

1. Regression
2. Binary Classification
3. Multiclass Classification



Supervised Learning Examples

Classification

- **Email Spam Filters**
  - Features: Words, sender, links
  - Label: Spam or Not Spam
- **Face ID / Fingerprint Unlock**
  - Features: Facial/fingerprint data
  - Label: You or Not You
- **Letter Recognition**
  - Features: Pixel values of images
  - Label: A-Z, 0-9, etc.

Regression

- **Weather Forecasting**
  - Features: Pressure, humidity, wind
  - Label: Temperature/rainfall amount
- **Uber/Lyft Pricing**
  - Features: Distance, time, demand
  - Label: Trip cost
- **YouTube View Count Predictions**
  - Features: Title, thumbnail, creator stats
  - Label: Expected views

Real-World Example

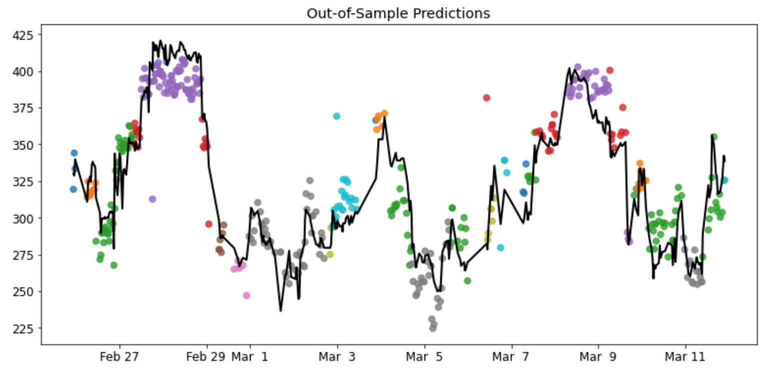
Outer Wall Thickness of an extruded vinyl profile is measured and recorded manually using a cut profile and a pair of calipers once every 12 hours.

- If this wall is **too thin**, the profile will create failure points.
- If this wall is **too thick**, the profile will be too heavy and expensive.
- Customer is **losing $400K/mo** in overage.

Metrics

- **Main Feed Speed**: Speed at which substrate material feeds into the main extruder.
- **Main Drive Amps**: Electrical load on the main extruder's motor.
- **Puller Speed**: Speed at which the product is pulled from the die.
- **Die and Barrel Zone Temperatures**: Temperature at the die and extruder barrel.

Dataset	Accuracy (>90%)	Correlation (>70%)	P95 error (<3 sigma)
Within-Sample	96%	94%	1.99 sigma
Out-of-Sample	95%	93%	1.73 sigma



Model

- Lasso Regression scored against a 60-minute rolling window
- Pre-aggregated data into 5-minute buckets

Exercise: Predicting Passengers

bigd103.link/linear-regression

Supervised Classification

Exercise

On average, which of the following Pokemon cards does the most damage? Write your answer in chat.

A Brief Interlude into Probability

Expected Value

Expected Value is a way to measure the average outcome of a random variable.

For example, “What’s the expected value of a standard die roll?” you’d add its possible outcomes and divide by 6:

$$\mathbb{E}[\text{die}] = \frac{1+2+3+4+5+6}{6} = 3.5$$

Using this, we can answer:

$$\begin{aligned}\mathbb{E}[\text{Quick Attack}] &= \left(\frac{1}{2}\right) \times 30 + \left(\frac{1}{2}\right) \times 10 \\ &= 15 + 5 \\ &= 20\end{aligned}$$

Multiplication Rule

The probability of something happening is denoted as:

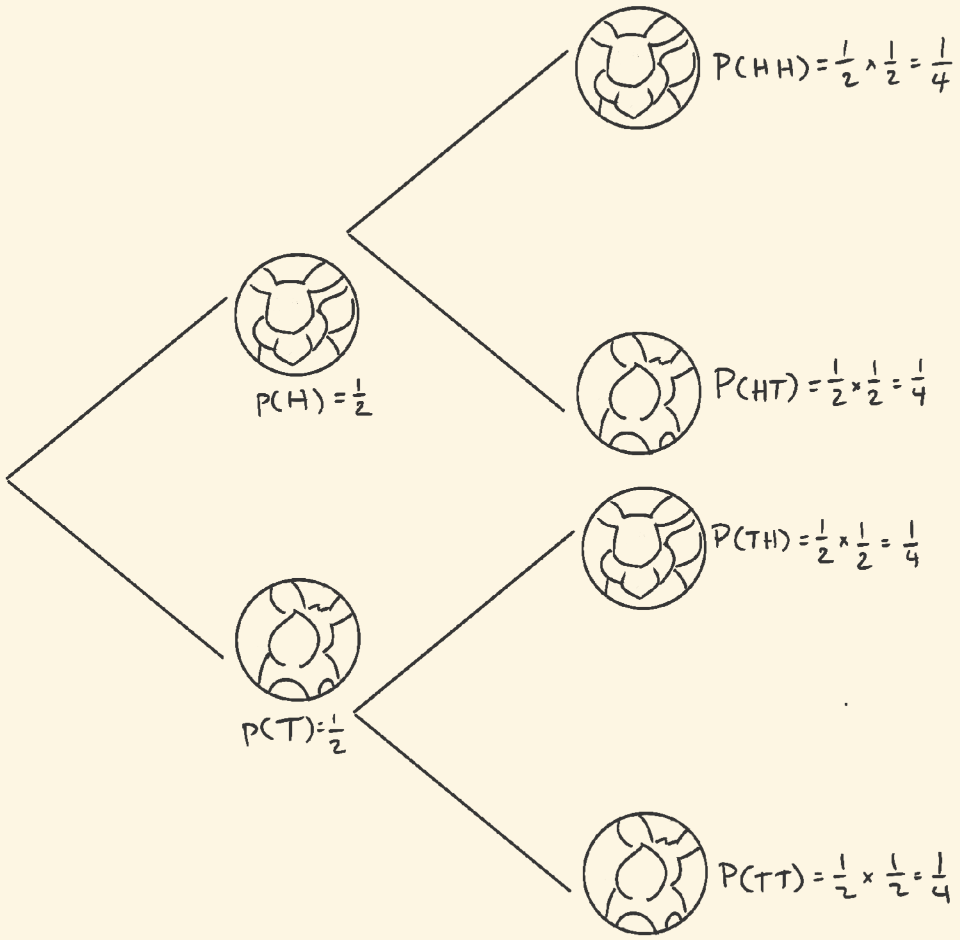
$$P(\text{event})$$

Two independent events have a joint probability:

$$P(A \cap B) = P(A) \times P(B)$$

For example, if the probability of flipping a coin on heads once is 1/2 than the probability of flipping it twice and getting heads both times is:

$$P(\text{heads}) \times P(\text{heads}) = 1/2 \times 1/2 = 1/4$$



Logistic Regression

Decision Boundary

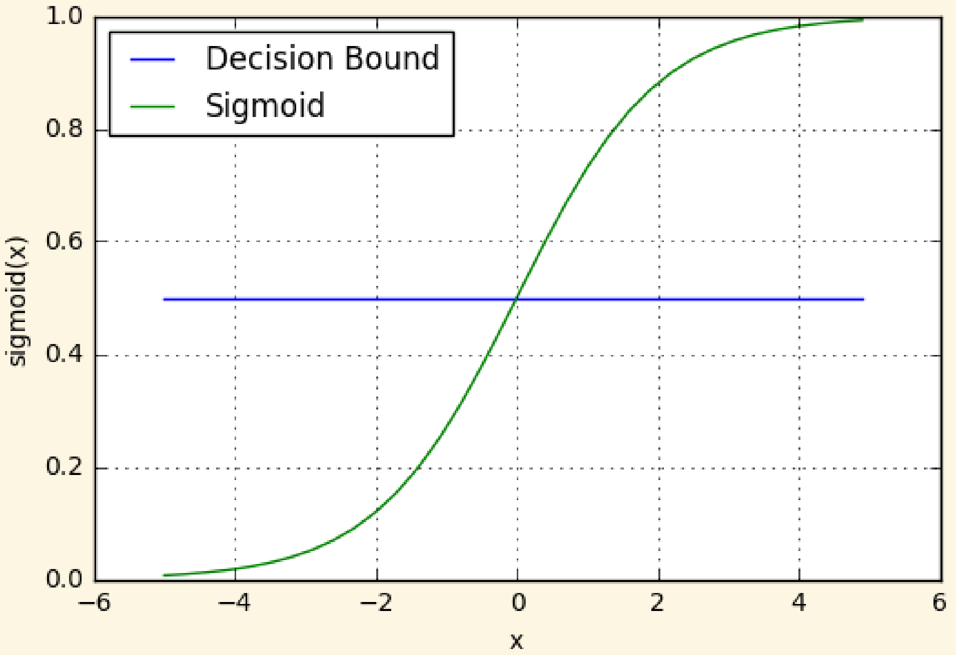
The decision boundary is defined where the model predicts $P(y = 1 \mid x) = 0.5$.

Equivalently, $\sigma(z) = 0.5 \rightarrow z = 0$.

Loss Function and Optimization

Uses the **Cross-Entropy Loss** (also called Log Loss).

Parameters (w_i) are optimized using **gradient descent** to minimize this loss.



Training a Supervised Learning Model

What is an SVM?

Support Vector Machines are powerful classifiers that find the **optimal decision boundary** between classes.

Key idea: Don't just find any line that separates the classes - find the line with the **maximum margin**.

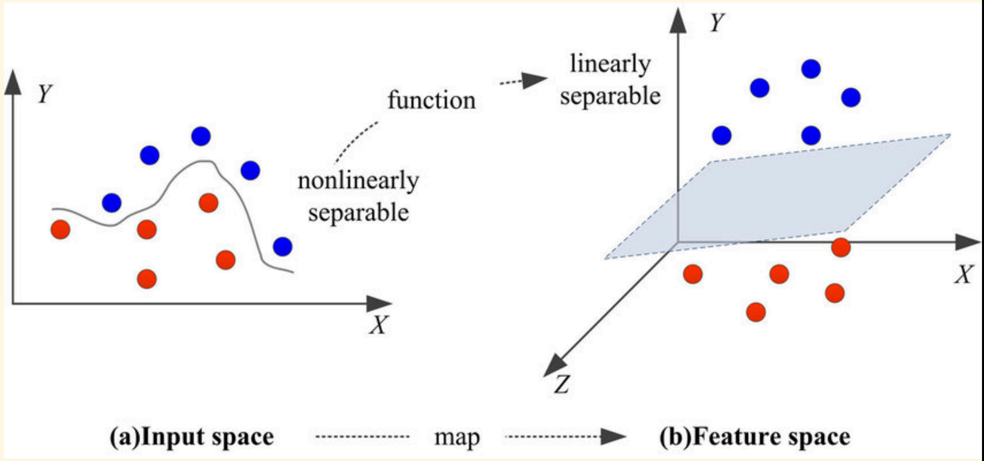
Margin: The distance between the decision boundary and the nearest data points from each class.

The Kernel Trick

Problem: What if the data isn't linearly separable?

Solution: The kernel trick allows SVMs to:

1. Map data to a higher-dimensional space
2. Find a linear boundary in that space
3. Which appears non-linear in the original space!



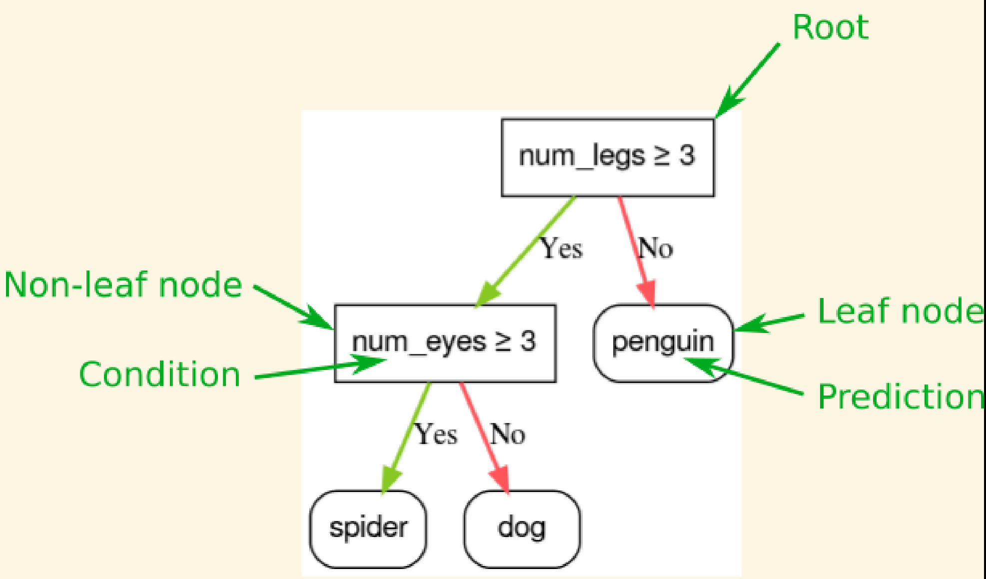
Common Kernels

- **Linear**: no transformation
- **RBF**: Radial Basis Function
- **Polynomial**: polynomial transformation

What is a Decision Tree?

A decision tree is a flowchart-like structure used for classification and regression tasks.

It recursively splits the dataset into subsets based on feature values, forming a tree of decisions.



Comparing Classifiers

Algorithm	Pros	Cons
Logistic Regression	Fast, probabilistic, interpretable	Linear only
SVM	Powerful, kernels, works in high-D	Slow, hard to interpret
Decision Trees	Interpretable, non-linear, fast	Can overfit easily

In practice: Try multiple algorithms and see what works best for your data and problem!

Machine Learning in Production