NBA 4920/6921 Lecture 1

Introduction to Machine Learning

Murat Unal

Johnson Graduate School of Management

08/31/2021

Agenda

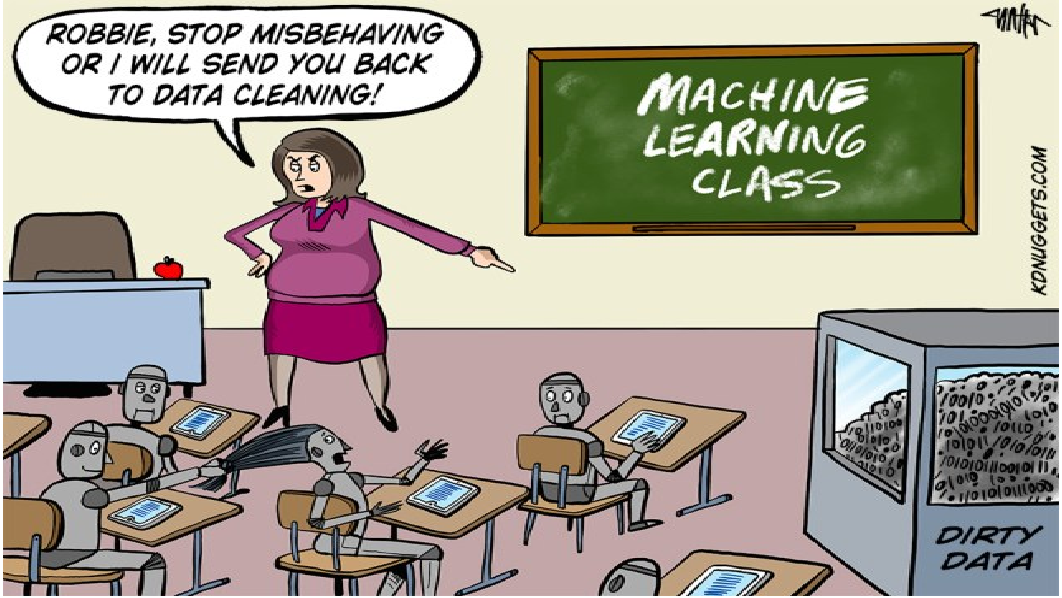
What is machine learning

Classes of machine learning

- Supervised learning
- Unsupervised learning

Type of models for machine learning

- Parametric models
- Non-parametric models



Source: kdnuggets.com

What is machine learning?

It is the process of solving a practical problem by

1. gathering a dataset
2. algorithmically building a statistical model based on the dataset.

It is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world.

What is machine learning?

It is the process of solving a practical problem by

1. gathering a dataset
2. algorithmically building a statistical model based on the dataset.

It is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world.

It's a term coined by Arthur Samuel, an IBM employee, in 1959 for marketing purposes.

What is machine learning?

- ML is so pervasive today that it is being utilized in many sectors and industries
- Anything you can imagine to be **predictable** is a candidate setting for applying ML
- Marketing: churn prediction, recommendations, advertising, pricing strategies, demand forecasting, improving customer service
- Healthcare: cancer detection on mammograms, identifying surgery candidates
- Financial services: fraud detection, credit scoring
- Education: which teacher will have the greatest value added
- Public policy: targeting health inspections
- Social policy: predicting highest risk youth for targeting interventions

Types of machine learning

Two main classes: **supervised learning** and **unsupervised learning**.

Supervised learning

Starts with gathering the data: collection of pairs (input, output)

Converts the pairs into machine-readable data

Uses input/feature/predictor/explanatory/independent variables represented by $X = (x_1, x_2, \cdots, x_p)$

to learn a function f for predicting an output/target/response/dependent variable represented by Y

Supervised learning

The goal is to build a model that captures the relationship between Y and X using a function f

$$Y = f(X)$$

$\rightsquigarrow$ Regression (linear, logistic, trees)

Unsupervised learning

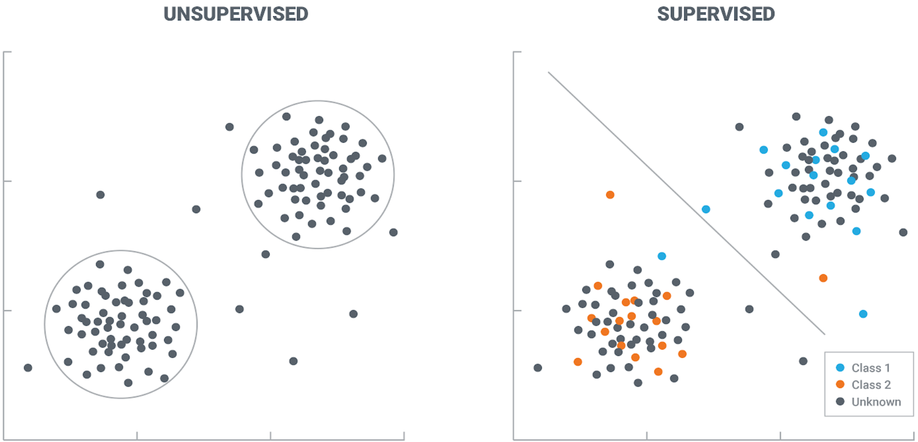
Uses only inputs X and no associated or *supervising* outputs Y to uncover relationships from the data

Lets the data *speak* for itself

Can also be described as *pattern recognition* algorithms

$\rightsquigarrow$ Clustering

Unsupervised vs. supervised learning



Source: lawtomated.com

Supervised learning

Is further separated into two groups depending on the characteristic of the output variable:

1. **Regression tasks:** Deal with *quantitative/numeric/continuous* outputs
   $\rightsquigarrow$ Income, price, sales etc.

2. **Classification tasks:** Deal with *qualitative/categorical/discrete* outputs
   $\rightsquigarrow$ loan default, digits (0-9), race, sex etc.

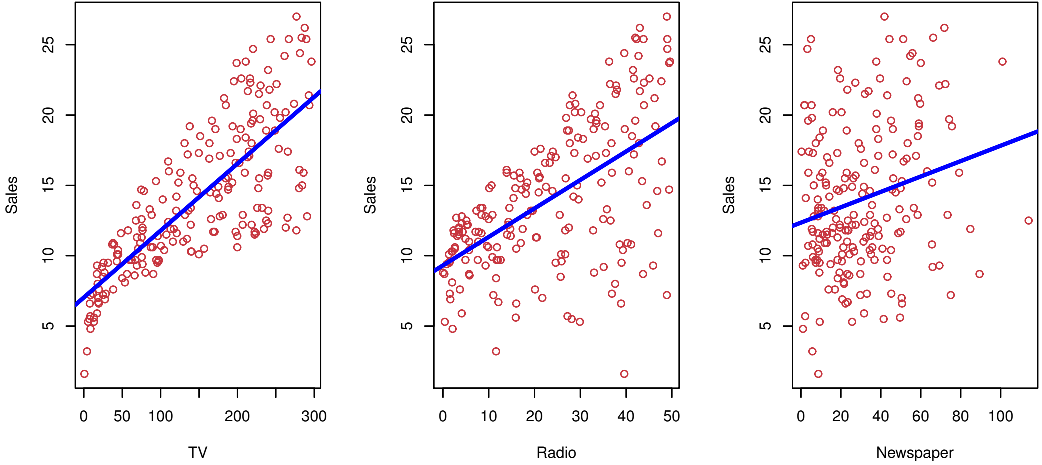
Statistical learning vs machine learning

- Statistical learning refers to a set of tools developed to understand/model data
- Both deal with supervised and unsupervised learning
- Machine learning was developed by computer scientists
- In ML emphasis is on large scale applications, i.e. high dimensional data, and **prediction accuracy**
- Statistical learning was developed by statisticians
- In SL emphasis is on models and their **interpretability** as well as **uncertainty**

Statistical learning vs machine learning

- "ML has the upper hand in Marketing!"
- What's the role of "learning"?
- Methods/algorithms **tune model parameters** based on the observed dataset we feed them, i.e they learn from the data thereby optimize parameters

Machine learning model



Source: ISL

Machine learning model

Goal: Build a model to understand `Sales` as a function of advertisement spent in different media.

Output/Target/Response/Dependent Variable:
Y = `Sales`

Input/Feature/Predictor/Explanatory/Independent Variable:
X = (`TV`,`Radio`,`Newspaper`)

Machine learning model

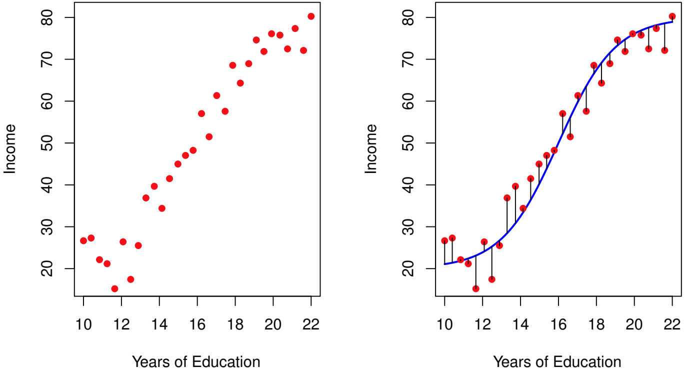
The relationship between output Y and p inputs, $X = (X_1, \cdots, X_p)$, can be written as

$$Y = f(X) + \epsilon$$

- f is an unknown function we want to learn/estimate
- It represents the **systematic** information that X provides about Y
- ϵ is a mean-zero error term that is independent of the inputs
- It represents the **noise/randomness/unobservables** that can not be explained using X

Machine learning model

The blue curve is the true underlying relationship we want to learn



Source: ISL

What can we use $\hat{f}$ for?

$$\texttt{Sales} = \hat{f}(\texttt{TV}, \texttt{Radio}, \texttt{Newspaper})$$

Using the observed data we learn/estimate f and obtain $\hat{f}$ for two main purposes.

What can we use $\hat{f}$ for?

$$\texttt{Sales} = \hat{f}(\texttt{TV}, \texttt{Radio}, \texttt{Newspaper})$$

Using the observed data we learn/estimate f and obtain $\hat{f}$ for two main purposes.

1. **Inference**: Is higher advertising expenditure associated with higher sales? Which media contributes more?

What can we use $\hat{f}$ for?

$$\texttt{Sales} = \hat{f}(\texttt{TV}, \texttt{Radio}, \texttt{Newspaper})$$

Using the observed data we learn/estimate f and obtain $\hat{f}$ for two main purposes.

1. **Inference**: Is higher advertising expenditure associated with higher sales? Which media contributes more?

   Q: Can we make causal claims? Does advertising increase sales?

What can we use $\hat{f}$ for?

$$\texttt{Sales} = \hat{f}(\texttt{TV}, \texttt{Radio}, \texttt{Newspaper})$$

Using the observed data we learn/estimate f and obtain $\hat{f}$ for two main purposes.

1. **Inference**: Is higher advertising expenditure associated with higher sales? Which media contributes more?

   Q: Can we make causal claims? Does advertising increase sales?

   A: With observational studies *usually* we can not make causal claims.

   Association $\neq$ Causation.

   Econometrics is the field that studies methods for causal inference in observational settings.

What can we use $\hat{f}$ for?

$$\texttt{Sales} = \hat{f}(\texttt{TV}, \texttt{Radio}, \texttt{Newspaper})$$

Using the observed data we learn/estimate f and obtain $\hat{f}$ for two main purposes.

2 **Prediction**: Predict sales from advertising expenditure.

How do we estimate f

Assume we have observed a set of n different data points, which are called the **training data**.

We use these observations to train a statistical learning method how to estimate the unknown function f

i.e. we want to find a function $\hat{f}$ s.t. $Y \approx \hat{f}(X)$

Most methods for this task can be characterized as either: **parametric** or **non-parametric**

Parametric models

First assumes a functional form of f then uses the training data to train/fit the model.

$\rightsquigarrow$ The **linear model**: $f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$

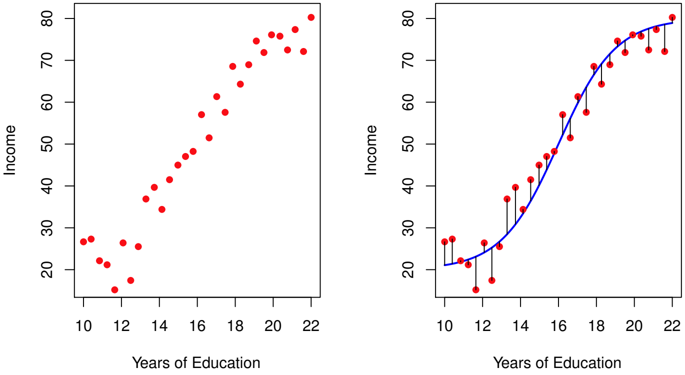
$\rightsquigarrow$ We can estimate the parameters $\beta_0, \beta_1, \beta_2, \cdots .\beta_p$ using **ordinary least squares** (OLS)

Pro: Easy to estimate

Con: Less flexible. Can be a poor approximation for the true unknown form of f

Parametric models

What do you think would happen if we estimated the true blue curve with a liner parametric model, such as OLS?



Source: ISL

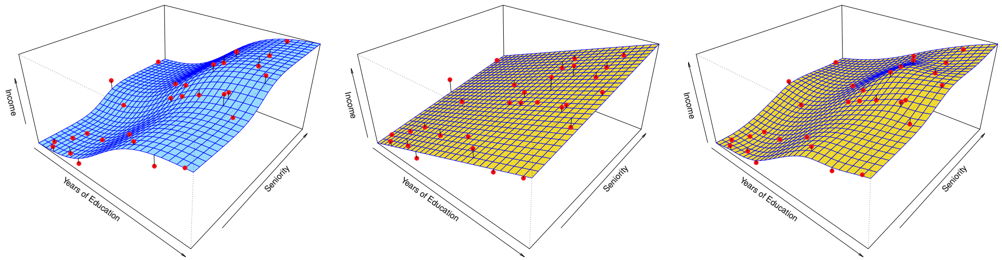
Non-parametric models

Do not make explicit assumptions about the functional form of f

$\rightsquigarrow$ regression trees, random forests

Pro: Increased flexibility. Can be a good approximation for the true unknown form of f

Con: Far more observations is required in order to obtain an accurate estimate for f



Source: ISL

1. True model: $\texttt{Income} = f(\texttt{Education} \ , \texttt{Seniority}) + \epsilon$
2. Parametric model: Linear regression fit to the data

   $\hat{f}(\texttt{Education} \ , \texttt{Seniority}) = \hat{\beta}_0 + \hat{\beta}_1 \texttt{Education} \ + \hat{\beta}_2 \texttt{Seniority}$
3. Non-parametric model: Spline fit to the data

Parametric or non-parametric

Previosly we mentioned that ML algorithms learn from the data thereby optimize model parameters.
Now we are saying one of the quintessential ML algorithm, i.e. random forests is a non-parametric model? Why?

Parametric or non-parametric

Previosly we mentioned that ML algorithms learn from the data thereby optimize model parameters.
Now we are saying one of the quintessential ML algorithm, i.e. random forests is a non-parametric model? Why?

Parametric models have parameters regarding the underlying model that generated the data. They make assumptions how the data was generated.

Whereas regression trees, RF or neural nets have parameters related with the algorithm itself, but they don't make assumptions about your data distribution or the underlying process that generated the data.

Next lecture

Data exploration and visualization.

Reminder: Wednesday (09/01) between 5:00-06:30 PM lab session 2 in Sage B08.

References

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2017)
An Introduction to Statistical Learning
Springer.
https://www.statlearning.com/

Andriy Burkov (2021)
The Hundred-Page Machine Learning Book
http://themlbook.com

Ed Rubin (2020)
Economics 524 (424): Prediction and Machine-Learning in Econometrics
Univ, of Oregon.