Applied Machine Learning (AML)

Introduction to Machine Learning

Oisin Mac Aodha • Siddharth N.

Why Machine Learning?

Machine Learning

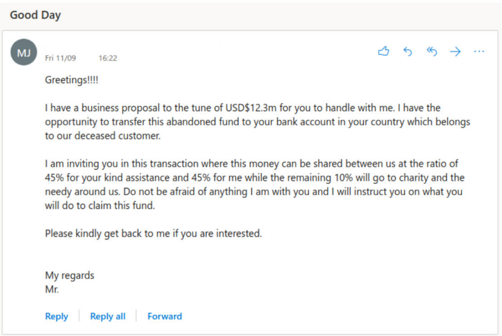
Machine Learning (ML) is the study and development of algorithms that **learn from data** in order to make **predictions** about new data.

$$y = f(\boldsymbol{x}; \boldsymbol{\theta})$$

f is our model
$\boldsymbol{\theta}$ are the parameters of our model
$\boldsymbol{x}$ are features
y is the model prediction

Example - Spam Detection

Can we determine if a new email we have received is **spam** or if it is important (**i.e. not spam**)?


Good Day

MJ Fri 11/09 16:22

Greetings!!!!

I have a business proposal to the tune of USD$12.3m for you to handle with me. I have the opportunity to transfer this abandoned fund to your bank account in your country which belongs to our deceased customer.

I am inviting you in this transaction where this money can be shared between us at the ratio of 45% for your kind assistance and 45% for me while the remaining 10% will go to charity and the needy around us. Do not be afraid of anything I am with you and I will instruct you on what you will do to claim this fund.

Please kindly get back to me if you are interested.

My regards
Mr.

Reply | Reply all | Forward

Example - Spam Detection

Can we determine if a new email we have received is **spam** or if it is something important (**not spam**)?

$$y = f(\boldsymbol{x}; \boldsymbol{\theta})$$

f is our model
$\boldsymbol{\theta}$ are the parameters of our model
$\boldsymbol{x}$ are features - a new email
y is the model prediction - spam / not spam

Why Learning

Manually defining exhaustive rules for a given task (e.g. spam detection) can be very challenging, as it is difficult to account for all possibilities.

Instead we can **learn** our model from a dataset, $\mathcal{D}_{train} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), ..., (\boldsymbol{x}_N, y_N)\}$.

$$\arg\min_{\boldsymbol{\theta}} = \frac{1}{N}\sum_{n=1}^{N} \ell(y_n, f(\boldsymbol{x}_n; \boldsymbol{\theta}))$$

Here, ℓ is a function that measures if the model's predictions are correct. For example, $\ell(.) = 0$ if the model is correct, and 1 otherwise.

Our goal is to choose the parameters $\boldsymbol{\theta}$ that make the fewest mistakes on the data.

Human Learning

Human learning is a process of acquiring knowledge and abilities. This results in a change in our cognitive structures due to from experience. This in turn changes the way we perceive.

Humans are highly effective learners, e.g. we can learn new tasks with limited prior exposure.

Our machine learning systems are loosely inspired by these processes, i.e. the goal is not necessarily to precisely model human learning.

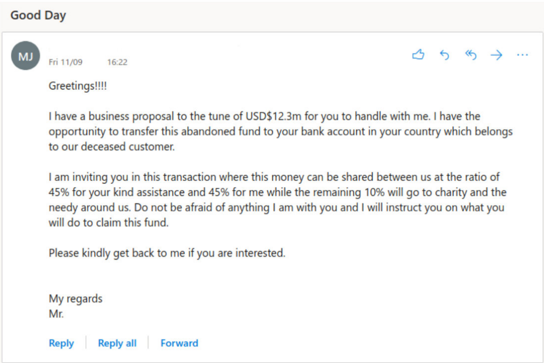
Core Questions

- What task am I trying to solve?
- How should I model the problem?
- How should I represent my data?
- How can I estimate the parameters of my model?
- How should I measure the performance of my model?

Example Machine Learning Tasks

Spam Detection

Is an email **spam** or **not spam**?


Good Day

MJ

Fri 11/09 16:22

Greetings!!!!

I have a business proposal to the tune of USD$12.3m for you to handle with me. I have the opportunity to transfer this abandoned fund to your bank account in your country which belongs to our deceased customer.

I am inviting you in this transaction where this money can be shared between us at the ratio of 45% for your kind assistance and 45% for me while the remaining 10% will go to charity and the needy around us. Do not be afraid of anything I am with you and I will instruct you on what you will do to claim this fund.

Please kindly get back to me if you are interested.

My regards
Mr.

Reply | Reply all | Forward

Image Classification

Determining what is present in an image, e.g. the species of animal.

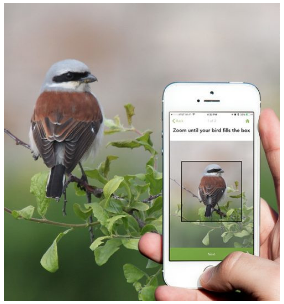


Image credit: https://ebird.org

Geospatial Prediction

Determining how many of a ‘thing’ we might expect to find in a region given features representing the local environment.

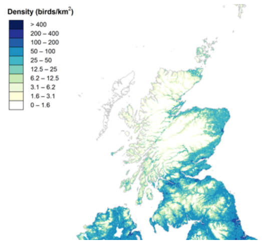


Image credit: https://www.bto.org/our-science/projects/breeding-bird-survey

Autonomous Driving

Determining the objects present in the scene and how to navigate safely around them.

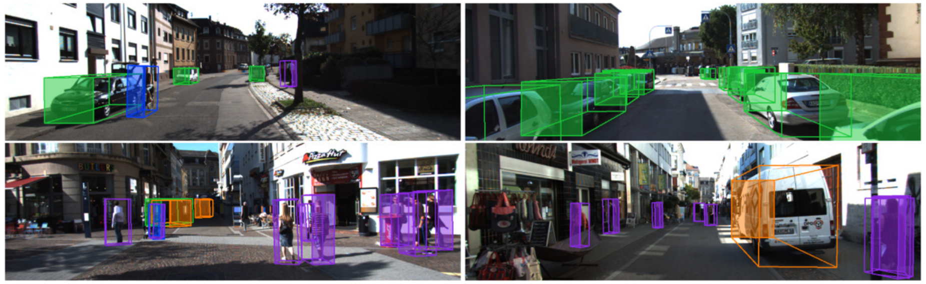
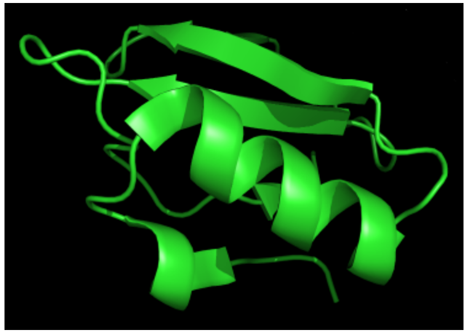


Image credit: https://arxiv.org/abs/2108.06417

Protein Folding

Determining a protein's 3D shape from the amino-acid sequence.



Recommender Systems

Predicting the preference of users for a set of items, e.g. what movie an individual may like to watch.

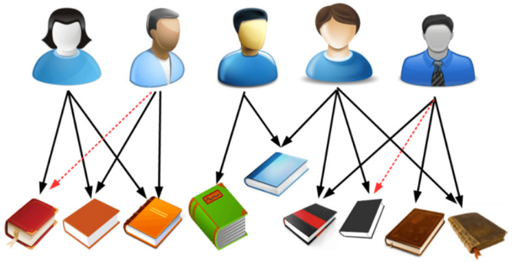


Image credit: https://thedatascientist.com/right-way-recommender-system-startup

Plus Many More!

- Healthcare
- Finance
- Robotics
- Sustainability
- Web Search
- Speech Recognition
- ...

With Great Power ...

Machine learning is a powerful tool that can be used for good or bad.

Often the ‘bad’ can stem from:

- Lack of care
- Ignoring potential biases in data
- Lack of understanding of model limitations
- Blindly trusting automated systems

Machine Learning Tasks

Supervised Learning

- The most commonly encountered form of machine learning is **supervised learning**
- Here the goal is to learn a function f from inputs (i.e. features) $\boldsymbol{x}$ to output y
- We learn this mapping using training data $\mathcal{D}_{train} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$
- Examples:
  - Classification: $y \in \{1, ..., C\}$
  - Regression: $y \in \mathbb{R}$

Unsupervised Learning

- In **unsupervised learning**, we only observe the features $\boldsymbol{x}$ without having access to any outputs
- As a result, the training data is now $\mathcal{D}_{train} = \{(\boldsymbol{x}_n)\}_{n=1}^{N}$
- Examples:
  - Clustering
  - Dimensionality reduction

Reinforcement Learning

- Here we are concerned with a system (i.e. agent) that needs to learn how to interact with its environment
- Our goal is to learn a policy which specifies the action to take in response to observation of the environment
- The agents aim is to maximise their reward

Acknowledgements

The slides in this course build on previous material from the *Introductory Applied Machine Learning* (IAML) course at the University of Edinburgh.

Credit to:

- Chris Williams
- Victor Lavrenko
- Charles Sutton
- Nigel Goddard
- and others