THE UNIVERSITY of EDINBURGH
informatics

Applied Machine Learning (AML)

Class Starting at 4:10pm

Oisin Mac Aodha • Siddharth N.

Applied Machine Learning

Week 2: Intro to ML and Classification

This slides will be made available on the project website after the class.
This session will be recorded.

Course Instructors

+ a big team including TA, lab demonstrators, and tutors helping out

Overview

1) Discussion of Week 1’s topics

2) Outline of your tasks this for Week 2

Course Website - **https://tinyurl.com/aml2024**

https://groups.inf.ed.ac.uk/teaching/aml/

Applied Machine Learning (AML)

Search Applied Machine Learning (AML)

DRPS | Piazza | Learn | Labs | Tutorials

- **Home**
- Course Information
- Schedule
- Mini-Project
- Exam
- Staff

Applied Machine Learning (MSc)

Informatics (INFR11211), Semester 1, 2024

Oisin Mac Aodha
oisin.macaodha@ed.ac.uk

Siddharth N.
n.siddharth@ed.ac.uk

Week 0 Announcement

Sep 10 · 0 min read

Welcome Week!

Announcements

*URL also available on **Learn***

Make sure to refresh the page

Q&A Sessions

Q&A sessions will be recorded so you can watch them offline later

Click here to find them on Learn - after some delay e.g. 1 day

Have you watched the lectures from week 1?

Lab 0: Introduction

Lab 0 “**00 - Introduction.ipynb**” in GitLab page

Getting setup on Notable - online Jupyter Notebook. We only support Notable

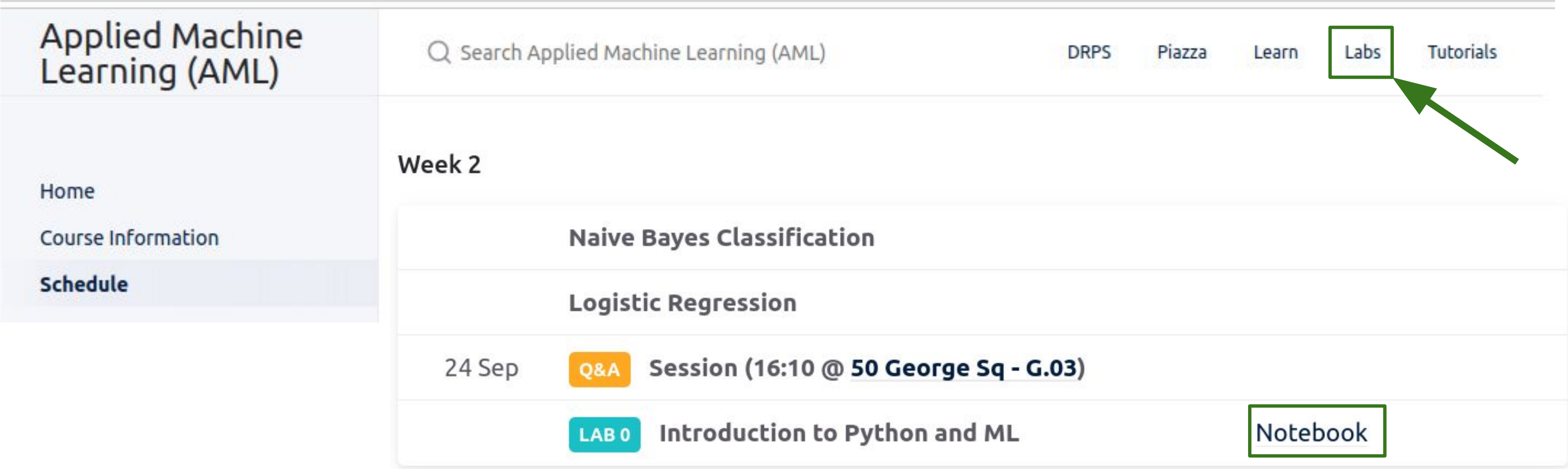
Introduction to Python and some of the core libraries we will use e.g. numpy

Important to be comfortably with this so that you can do the labs and courseworks in future weeks. Get started now!

Lab 0: Introduction

URL for Lab 0 is available on the course webpage -> schedule

Links to solutions will also be provided 1 week later



Have you completed Lab 0?

Optional - Drop in Lab

Did you have an issues with Lab 0?

If so, we have a drop in lab session from **1-3pm** in **Appleton Tower 4.12** tomorrow (25th Sep)

No need to attend if you completed most of Lab 0. This is only for people who are stuck.

Note, if you found Lab 0 very difficult, consider if AML is a good fit for you.

Lab and Tutorial Groups

Your Lab groups assignment will be performed by ITO/Timetabling

Labs: There are twelve lab sessions per week LAB01:LAB12 - **starting next week**

Tutorials: There will only be 7 (i.e. not 8) tutorials. Fri at 1:10 will not happen.

Only go to one of them, i.e. the one you have been assigned to

Note, check the time of your lab/tutorial in advance as one of the labs has moved i.e. Lab12 is on Tues at 10am (moved from Monday at 10am)

Course Selection Deadline

- The deadline for Semester 1 course changes is the end of week 2 of Semester 1
- If you wish to make any course changes for Semester 1, please email your Student Adviser as soon as possible, and before the end of week 2
- After this date we cannot change your Semester 1 courses

https://web.inf.ed.ac.uk/infweb/student-services/taught-students/information-for-students/information-for-msc-students/taught-msc-handbook-2024-25/registration-change

Lecture Review

- Introduction to machine learning tasks

- Showed how simple statistical models, e.g. multivariate Gaussians, can be used to perform tasks such as classification
- Showed how we can fit these models to data use maximum likelihood estimation
  - i.e. how to estimate the parameters of the models from data

Introduction to Machine Learning

- Examples of machine learning
- Different machine learning tasks, e.g. classification, regression, …

Classification

- In *classification* we are assigning objects/concepts to a set of **pre-existing** classes/categories

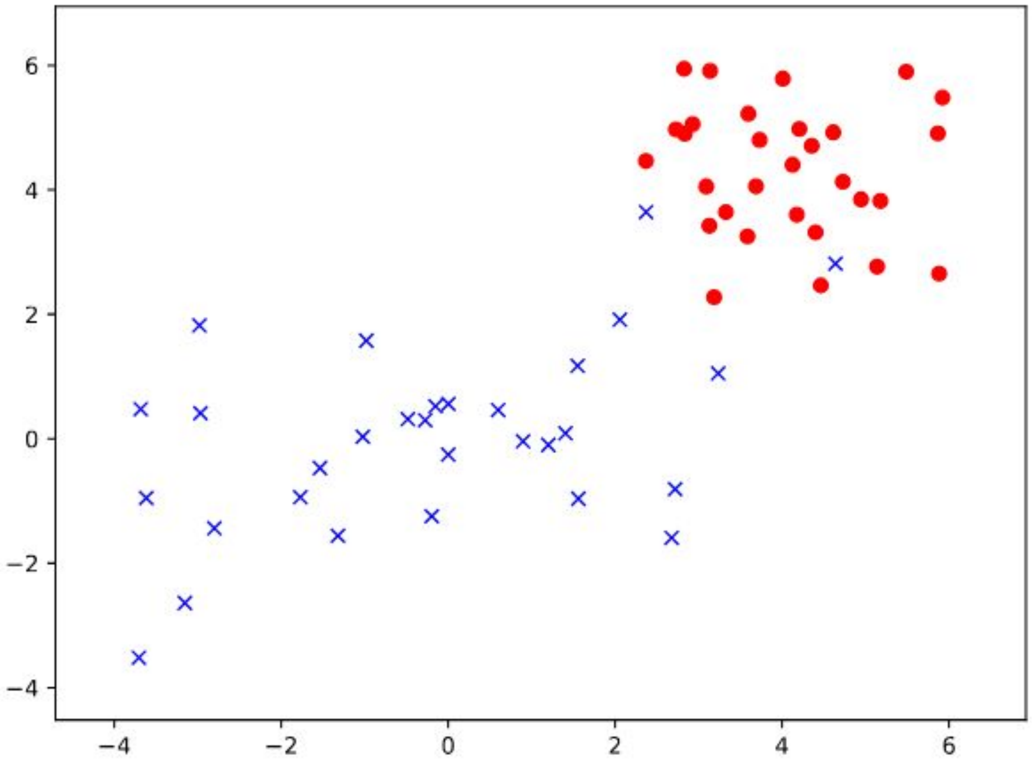
- Can be a good shortcut to quickly assign objects/concepts to categories, e.g. safe / poisonous
  - Good to have mental shortcuts for making decisions quickly

Classification

“Chihuahua Or Muffin?”

Source A. Gri, "Puppies Or Food? 12 Pics That Will Make You Question Reality", Bored Panda, 2022.

Classification



$$p(y = c|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|y = c)p(y = c)}{\sum_{c'} p(\boldsymbol{x}|y = c')p(y = c')}$$

$$p(\boldsymbol{x}|y = c) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_c, \Sigma_c)$$

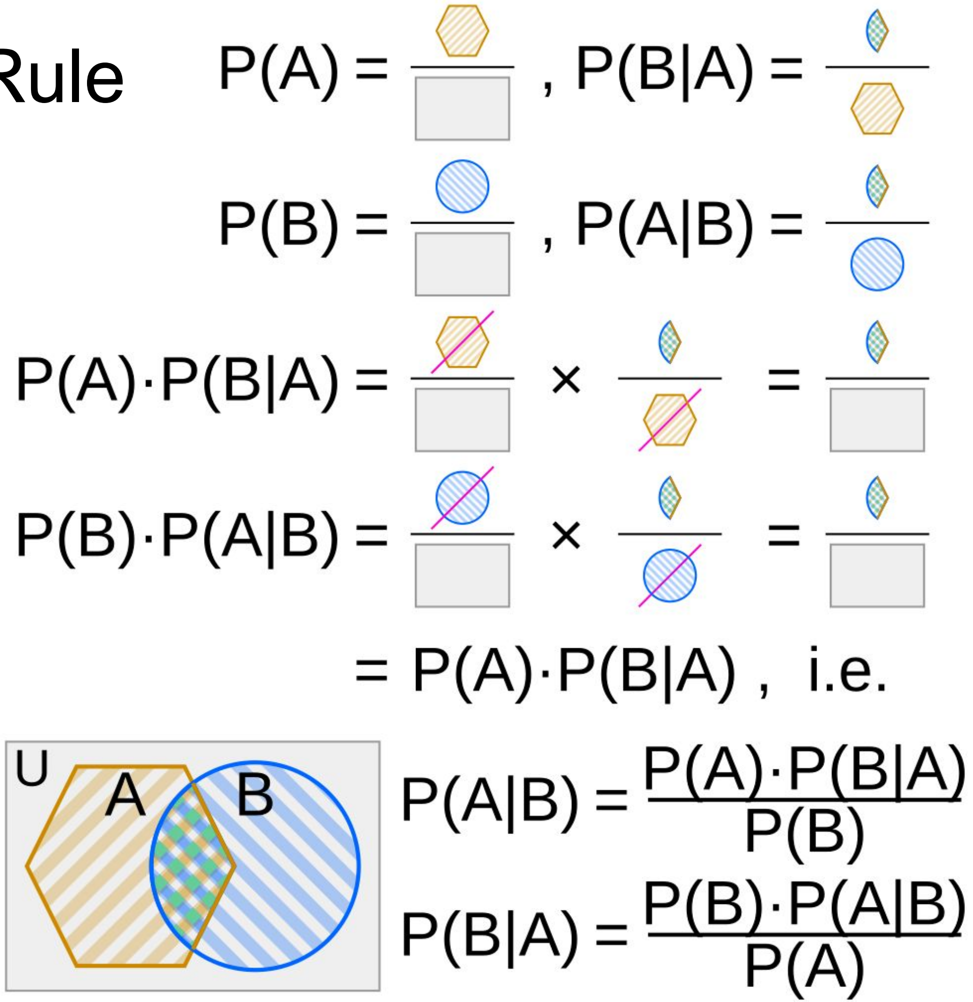
Week 2: Your tasks for this week

1) Complete Lab 0
    a) Attend the drop in lab if needed
2) Watch the videos for week 2 - Naive Bayes and Logistic Regression
    a) Ask questions on Piazza if stuck
3) Start Lab 1 - link in week 3

Week 2

Naive Bayes Classification	Playlist • Slides • Handout
Logistic Regression	Playlist • Slides • Handout

Visual “proof” of Bayes Rule



Source
https://en.wikipedia.org/wiki/Bayes%27_theorem

Maximum Likelihood for Bernoulli Distribution

Given a dataset $\mathcal{D} = \{(x_1), (x_2), ..., (x_N)\}$, where x_n is a binary variable (i.e., $x_n \in \{0, 1\}$), we would like to fit a model to our data that has parameters $\boldsymbol{\theta}$. We will need to estimate these parameters using the data.
We start be defining the likelihood of our data as:

$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{n=1}^{N} p(x_n|\boldsymbol{\theta})$$

Here we have assumed that the data is independent and identically distributed (IID). We are going to estimate the parameters using maximum likelihood estimation (MLE):

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\mathcal{D}|\boldsymbol{\theta})$$

In practice, this is equivalent to minimizing the negative log likelihood (NLL):

$$\begin{aligned}
\text{NLL}(\boldsymbol{\theta}) &= -\log\left[p(\mathcal{D}|\boldsymbol{\theta})\right] \\
&= -\log\left[\prod_{n=1}^{N} p(x_n|\boldsymbol{\theta})\right]
\end{aligned}$$

We can make use of the fact that the log of a product is the same as the sum of the logs, i.e., $\log(ab) = \log(a) + \log(b)$.

$$= -\sum_{n=1}^{N} \log \left[p(x_n|\boldsymbol{\theta})\right]$$

As the data is binary, a natural choice for our model is the Bernoulli distribution which has a single parameter (i.e., $\boldsymbol{\theta} = \{\phi\}$). Swapping in the expression for the Bernoulli, we can write our NLL as:

$$= -\sum_{n=1}^{N} \log \left[\phi^{x_n} (1-\phi)^{(1-x_n)}\right]$$

We can again employ the same log identity to separate out this expression into two terms:

$$= -\sum_{n=1}^{N} \left(\log \left[\phi^{x_n}\right] + \log \left[(1-\phi)^{(1-x_n)}\right]\right)$$

We can also use the fact that $\log(a^b) = b\log(a)$ to obtain:

$$= -\sum_{n=1}^{N} \left(x_n \log \left[\phi\right] + (1-x_n) \log \left[(1-\phi)\right]\right)$$

We can further simplify the notation by defining $N_1 = \sum_{n=1}^{N} x_n$, $N_0 = \sum_{n=1}^{N}(1-x_n)$, and $N = N_1+N_0$. In the case of N_1, this represents the count of the number of times that the variable x_n is positive in the dataset:

$$= -\left(N_1 \log[\phi] + N_0 \log[(1-\phi)]\right)$$

To obtain the maximum likelihood estimate (MLE) of the parameter of interest ϕ, we take the derivative of the NLL with respect to ϕ and set it to zero:

$$\frac{d}{d\phi}\text{NLL}(\phi) = -\frac{d}{d\phi}\left(N_1 \log[\phi] + N_0 \log[(1-\phi)]\right)$$

To obtain an expression for the derivative, we can make use of the fact that $\frac{d\log(f(a))}{da} = \frac{1}{f(a)}\frac{df(a)}{fa}$, i.e., $\frac{d\log(a)}{da} = \frac{1}{a}$ and $\frac{d\log(1-a)}{da} = -\frac{1}{1-a}$:

$$\frac{d}{d\phi}\text{NLL}(\phi) = -\frac{N_1}{\phi} + \frac{N_0}{(1-\phi)}$$

Next we set this to zero and rearrange, to find our MLE of ϕ, i.e., $\hat{\phi}$:

$$
\begin{aligned}
0 &= -\frac{N_1}{\phi} + \frac{N_0}{(1-\phi)} \\
0 &= -\frac{N_1(1-\phi)}{\phi(1-\phi)} + \frac{\phi N_0}{\phi(1-\phi)} \\
0 &= -N_1(1-\phi) + \phi N_0 \\
0 &= -N_1 + N_1\phi + \phi N_0 \\
0 &= -N_1 + \phi(N_1 + N_0) \\
0 &= -N_1 + \phi(N_1 + N_0) \\
\hat{\phi} &= \frac{N_1}{(N_1 + N_0)} = \frac{N_1}{N}
\end{aligned}
$$

Intuitively, this is the empirical frequency of x_n being positive in the dataset.