Advanced NLP

L1 - 04/08/22

What is ML? M = f(theta)

What is a model? Its a replica of something we have already seen. Borrowed from statistics (a mathematical description of...)

What are stochastic processes? [What is the probabilities of Intro to NLP ppl choosing this course?]

Image processing and Computer vision are two different topics. (As are NLP and NLU, but let's not draw parallels)

Just counting the examples is not enough [Given a movie comment stats by a certain actor, per se, can't help you predict the nature of comments on the same actor's any new movie as a greater source of variability is brought forth by content or the nature of the film.]

M = f(theta, X)

theta = hyperparameters, X = the entities themselves

How to describe these entities? X is an entity which is a movie of SRK. The description of entity depends on the needs of the user and the user given definition of the entity.

Which entity featured to choose? Is it complete? Some factors are more important/fundamental than others.

Hence, the entities are comprised of **a set of features**. The features could be anything (depends on the task)

Assumptions:

1. each model is related to some purpose

P(Success = 1/0 | features). This is a truly statistic model but it might be practically impossible to solve.

P(A|B) = [P(B|A)* P(A)]/P(B)

Therefore, P(S=1 | features) = [P(features|S=1)*P(S=1)]/P(features)

Now, I can assume that each feature is independent of each other.

So what we have done just now is describe the model on the basis of features.

What is decision tree? How does it work?

When you use entities which are so fundamentally different that they cannot be put into spectrum, we need decision trees. It also tries to find out the importance of each feature, and to determine, for success, the most important feature. In decision tree, we basically rank the features.

What is a latent variables? For the class, let's say, students come from different distributions. So in the case the latent variable is the program individuals are coming from. Why latent? If sir doesn't ask our age and the program we come from, he will probably have a very misleading notion or understanding of the distribution.

Why are we talking about ML in an NLP class? Because for true analysis, we cannot just depend on pure mathematics. Theta, the hyperparameter, in real life, is either created by observation OR we start off with a theta and adjust it to maximise the likelihood of the observation given theta [max theta L(O|theta)]

The model can only maximise the likelihood for Ot (Trained observations) i.e. observations the model has already seen.

Model architecture are our choice

No model is perfect

Architectures do not make for a good solution, the problems make for architectures. There is a general tendency to pick up a model because someone else had done it previously.

Logistics of the course

1. Project Component - rather extensive ~ 40-50%
2. Assignments - 20%
3. End Sem - 20-30%
4. Quiz - 10%
5. Paper presentation - 10% (maybe)