Lecture 6 | Support Vector Machines

Date de création	@January 7, 2025 4:17 PM
Étiquettes	

Questions

- What are the key ideas?
- What terms or ideas are new to me?
- How would I define them?
- How do the ideas relate to what I already know?
- Should not assume that the profesor are always correct. Asking appropriate questions.
- What are the good ideas?
- Do the ideias have other applications?

▼

- Generative models looks at each classe, one step at time.
- Descriminative models use Gradient Descent to find the best fit line.

> **Generative** models learn backwards (features to class).

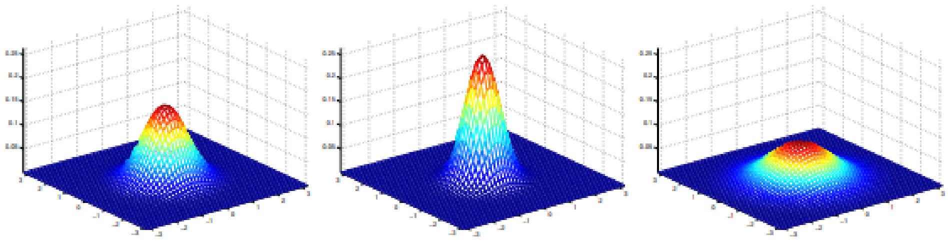
▼

This formula tell us how data behave, and we can use this to make sense of this behave and make cool stuff, like:

- predicti the future.
- analyse data.
- understand patterns.

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

Here are some examples of what the density of a Gaussian distribution looks like:



▼

Even they end up using the same Sigmoid Function GDA makes Strong Assumptions (About the Distribution Of Data) when Logistic Regression Don’t.

Using GDA means you know more things (in theory) that makes a good model, since you already knows those things.

Lecture

Outline

Naive Bayes

- Laplace Smoothing
- Event Models

Applications ML

SVM Intro

One simple problem when using Naive Bayes for text classification (spam/not spam) is that it can classifies something as no spam, just because the model never saw the word. Laplace smoothing helps adress this problem.

Seems like Laplace Smoothing consists in add 1 to the nominator and one to the denominator. Basicaly because you want to elimite the occurance of zero.

Some models has limitations that we can improve by using some mathematical trick.

Multinominal Event Model is a more efficient adaption of Naive Bayes focused on text classification, we create a much lower matrix in comparision with Multivariate Bernoulli Event Model. The Multinomial use a product with length equals to the number of words in the e-mail (for exemple) and instead of using {0, 1} we use {1, 2.., Num Words}. To apply Laplace we add 1 to the numerator and the legth of email to the denominator.

Even that, for most cases Logistic Regression is superior to Naives Bayes in accuracy, ther are some advantages in using it:

- simple to implement
- don’t use gradient
- computacionally efficient

SVM is great when you want classify data using a non-linear boudary. We can apply the same logic to Logistic Regression (non-linear boudary) by augmenting data. But, now how create this situation could be hard, so what SMV do is just create this high dimenstional space and then apply I linear classifier.

Two basic ideia in SMV are: Optimal Margin Classifier and Kernels. Kernels consist in convert feature to much higher dimensions.

Optimal margin: choose w and b that maximizes the Geometric Margin Denoted By γ.

Vocab

- Laplace smoothing: Laplace smoothing is a smoothing technique that handles the problem of zero probability in Naïve Bayes.
- NIPS: **Conference on Neural Information Processing Systems**
- Functional Margin:The functional margin is a measure of the generalization ability of the classifier.
- Geometric Margin: The distance is between the point and the decision boundary is Geometric Margin
- Kernel: It is a mathematical function that helps organize data and make it easier to classify.

More

- https://towardsdatascience.com/laplace-smoothing-in-na%C3%AFve-bayes-algorithm-9c237a8bdece
- https://neurips.cc/
- https://medium.com/@gokcenazakyol/machine-learning-support-vector-machines-svms-7bc389c877d8
- https://medium.com/@abhishekjainindore24/svm-kernels-and-its-type-dfc3d5f2dcd8
- https://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf