Lecture 5 | GDA and Naive Bayes

Date de création	@January 5, 2025 9:39 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?

▼ What's the difference between **Generative** and **Discriminative** models?

- 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).

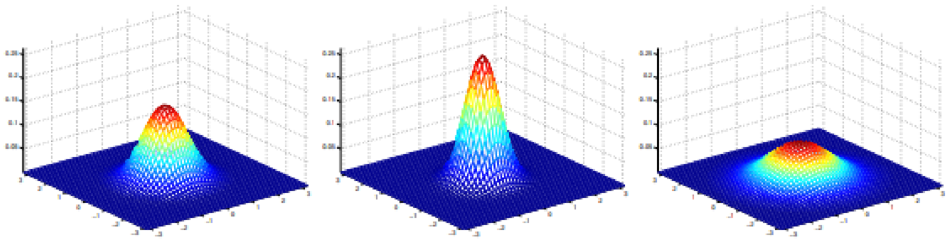
▼ Explain to a five year old kid, what this formula tells and why does it matter (Multivariate normal distribution).

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:



▼ When makes sense using **Generative** and When makes using **Descriminative**?

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

This week we are studying "generative models" in oposition to discriminative models (focus on predictions).

- Gaussian Descriminative Analysis (GDA)
- Generative vs Descriminative Composition
- Naive Bayes (Use for build a spam filter)

When working with ML Models I see we're constantly assuming that the **distribution is normal.** So, it's quite important, I understand, comprehend about the **normal distribution.**

Another word I see again and again in the classes are: **mean** and **variance**. Seems like basic concepts that have huge impact in ML.

It seems like GDA draws Gaussians and try predict the class (new one) using Baye's Rule.

GDA and Logit ends using Sigmoid Function to classify new data.

Certain models depends you know or make strong assumptions about data.

To use let's say **Naive Bayes** for e-mail classification we need, first, decide how to transform our text in a **feature vector**. One solution might by create a binary vector, every time a specific word appears we assign 1. It won't work, because it means 10k parameters. How about supose that: **Assume X's are conditionally independent given y.**

Models how only counts as Naive Bayes and GDA are more efficient that interative models like Logistic Regression.

Vocab

- Bayes Rule: ***Bayes' theorem*** gives a mathematical rule for inverting conditional probabilities, allowing one to find the probability of a cause given its effect.
- Multivariate normal distribution: The ***multivariate normal distribution*** is often used to describe, at least approximately, any set of (possibly) correlated real-valued random variables.
- Gaussian distribution: A Gaussian distribution, also known as a normal distribution, is a continuous probability distribution characterized by its bell-shaped curve.
- GDA model: Gaussian Discriminant Analysis (***GDA***) is a supervised learning algorithm used for classification tasks in machine learning.

- Joint Likehood: is the probability distribution of all possible pairs of outputs of two random variables that are defined on the same probability space.
- Covariance Matrix: a covariance matrix is a ***square matrix giving the covariance between each pair of elements of a given random vector***.
- Maximum Likelihood Estimation: Maximum likelihood estimation is ***a method that determines values for the parameters of a model***.
- Argmax: The *argmax* function returns the argument or arguments (*arg*) for the target function that returns the maximum (*max*) value from the target function.
- Naive Bayes: In statistics, **naive Bayes classifiers** are a family of linear "probabilistic classifiers" which assumes that the features are conditionally independent, given the target class.
- Conditionally Independent: **the situation where a variable is independent of another set of variables given a specific set of variables**
- Chain Rule (Probability): We can rearrange the formula for conditional probability to get the so-called product rule: ***P(A,B) = p(A|B) p(B)***.
- Probabilistic Graph Model: A **graphical model** or **probabilistic graphical model** (**PGM**) or **structured probabilistic model** is a probabilistic model for which a graph expresses the conditional dependence structure between random variables.

More

- https://en.wikipedia.org/wiki/Bayes%27_theorem
- https://en.wikipedia.org/wiki/Multivariate_normal_distribution
- https://www.datacamp.com/tutorial/gaussian-distribution
- https://www.geeksforgeeks.org/gaussian-discriminant-analysis/
- https://en.wikipedia.org/wiki/Joint_probability_distribution
- https://en.wikipedia.org/wiki/Covariance_matrix

- https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1
- https://machinelearningmastery.com/argmax-in-machine-learning/
- https://en.wikipedia.org/wiki/Naive_Bayes_classifier
- https://en.wikipedia.org/wiki/Chain_rule_(probability)
- https://en.wikipedia.org/wiki/Graphical_model