Machine Learning
Indian Institute of Technology Jammu
Generative and Discriminative Classifiers

Q1: In a Gaussian Naïve Bayes classifier, we typically assume that the standard deviation σ_i of $P(X_i|Y=k)$ is the same for all class values k. Now, suppose we relax this assumption and allow σ_{ik} to vary depending on both the feature index i and the class k. Given that $P(X_i–Y=k)$ follows a Gaussian distribution $N(\mu_{ik}, \sigma_{ik})$, derive the new form of $P(Y–X)$. Does this modified Naïve Bayes classifier still result in a logistic regression-like form for $P(Y–X)$? Justify your answer with a derivation.

Q2: Consider a Gaussian Bayes classifier where Y is a Boolean variable following a Bernoulli distribution with probability $\pi = P(Y=1)$. Instead of assuming independence between features, we now consider two continuous attributes, X_1 and X_2, which follow a bivariate Gaussian distribution $N(\mu_{1k}, \mu_{2k}, \sigma_1, \sigma_2, \rho)$ given $Y=k$. The means μ_{1k} and μ_{2k} depend on Y, but the standard deviations σ_1, σ_2 and correlation ρ remain constant across Y.

Using the given probability density function of the bivariate Gaussian distribution, derive the form of $P(Y|X)$. Does the resulting expression resemble the form used in logistic regression? Justify your answer with mathematical derivation.

Q3: Suppose you are training a generative naive Bayes model using a dataset where the distribution of labels does not accurately represent real-world conditions or the distribution in the test data. Given this scenario, which probability should you use to make the best decision on whether to predict y_0 for a given input X?

Q4: How do graphical representations of generative models, such as mixture models, directed graphical models, and undirected graphical models (Markov Random Fields), differ in their structure and applications? Provide examples of their use in real-world scenarios.

Q5: How does a generative classifier based on Bayes' theorem utilize the joint probability distribution, and what computational challenges arise when estimating probabilities for high-dimensional binary variables?

Q6: Describe the architecture of a Generative Adversarial Network and how the generator and discriminator interact during training.

Q7: How do conditional generative models differ from unconditional ones? Provide an example scenario where a conditional approach is beneficial.

Q8: How does overfitting manifest in generative models, and what techniques can be used to prevent it during training?

Q9: What is gradient clipping, and how does it help in stabilizing the training process of generative models?

Q10: Discuss strategies for training generative models when the available dataset is limited.

Q11: Describe the concept of learning rate scheduling and its role in optimizing the training process of generative models over time.

Q12: Compare and contrast the use of L1 and L2 loss functions in the context of generative models. When might one be preferred over the other?

Q13: How do discriminative and generative models differ in their learning approach, what are their advantages, and in what scenarios would each be preferred in deep learning applications?

Q14: How does Bayes's theorem ($P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$) establish the relationship between generative and discriminative models?

Q15: What are the key probabilistic differences between generative models ($P(X,Y) or P(X)$) and discriminative models ($P(Y|X)$)?

Q16: Imagine you are a language classification system.

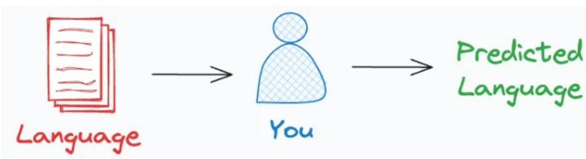

Figure 1: Language Classification System.

There are two ways you can classify languages.

a. Learn every language and then classify a new language based on acquired knowledge.

b. Understand some distinctive patterns in each language without truely learning the language. Once you do that, classify a new language based on the learned patterns.

Can you figure out which of the above is generative and which is discriminative?
Q17: What happens when we give correlated features in generative and discriminative models?
Q18: What happens when training data is biased over one class in generative and discriminative models?
Q19: How do generative and discriminative classifiers differ in terms of the probability distributions they model and the assumptions they make during training?