Title of Document

Name of Author

April 17, 2023

Latent variable modeling is a statistical framework that involves modeling observed data as the result of underlying variables that are not directly observable. These unobserved variables are known as latent variables and they are assumed to explain the variation in the observed data. One popular approach for modeling latent variables is through the use of probabilistic graphical models, which are graphical representations of how variables relate to each other probabilistically.

In many cases, it is difficult to directly infer the values of these latent variables from the observed data. Therefore, we use a technique called variational inference to approximate the posterior distribution of the latent variables given the observed data. Variational inference involves defining a family of distributions over the latent variables and then finding the member of that family that best approximates the true posterior distribution. This is done by minimizing a measure of the difference between the true posterior and the approximating distribution, which is known as the evidence lower bound (ELBO).

The ELBO is a lower bound on the log marginal likelihood of the data, and it can be expressed as:

$$\mathcal{L}(\theta, \phi; \mathbf{X}) = Eq\phi(\mathbf{Z}|\mathbf{X})[\log p_\theta(\mathbf{X}|\mathbf{Z})] - \mathrm{KL}(q_\phi(\mathbf{Z}|\mathbf{X}); |; p(\mathbf{Z}))$$

where $\mathbf{X}$ is the observed data, $\mathbf{Z}$ is the vector of latent variables, $p_\theta(\mathbf{X}|\mathbf{Z})$ is the likelihood function, $q_\phi(\mathbf{Z}|\mathbf{X})$ is the variational distribution over the latent variables, $p(\mathbf{Z})$ is the prior distribution over the latent variables, θ and ϕ are the parameters of the likelihood and variational distributions, respectively, and $\mathrm{KL}(q_\phi(\mathbf{Z}|\mathbf{X}); |; p(\mathbf{Z}))$ is the Kullback-Leibler (KL) divergence between the variational distribution and the prior.

The ELBO can be interpreted as the amount of information about the observed data that is explained by the latent variables minus the amount of information about the latent variables that is not explained by the observed data. The first term, $Eq\phi(\mathbf{Z}|\mathbf{X})[\log p_\theta(\mathbf{X}|\mathbf{Z})]$, is the expected log-likelihood of the observed data given the latent variables, and the second term, $\mathrm{KL}(q_\phi(\mathbf{Z}|\mathbf{X}); |; p(\mathbf{Z}))$, is the KL divergence between the variational distribution and the prior.

Jensen's inequality is a fundamental result in probability theory that states that for any convex function f, the expected value of $f(x)$ is greater than or equal to $f(E[x])$, i.e.,

$$E[f(x)] \geq f(E[x])$$

In the context of latent variable modeling, Jensen's inequality is used to derive a lower bound on the log-likelihood of the observed data. Specifically, we apply Jensen's inequality to the logarithm of the likelihood function, which is a concave function, to obtain:

$$\log p_\theta(\mathbf{X}) = \log \int p_\theta(\mathbf{X}, \mathbf{Z}); d\mathbf{Z} \geq \int q_\phi(\mathbf{Z}|\mathbf{X}) \log \frac{p_\theta(\mathbf{X}, \mathbf{Z})}{q_\phi(\mathbf{Z}|\mathbf{X})}; d\mathbf{Z} = \mathcal{L}(\theta, \phi; \mathbf{X})$$

where the inequality follows from Jensen's inequality, and the last expression is the ELBO.

Autoencoders are a type of neural network that can be used for unsupervised learning of latent representations of the observed data. An autoencoder consists of two parts: an encoder that maps the observed data to a latent representation, and a decoder that maps the latent representation back to the observed data. The objective of an autoencoder is to minimize the reconstruction error, which is typically measured by the mean squared error between the input and output.

To incorporate latent variables into an autoencoder, we can add a layer oflatent variables between the encoder and decoder. This layer is typically modeled as a probabilistic encoder that maps the observed data to a mean and variance of a Gaussian distribution in the latent space. The decoder is then modeled as a conditional distribution over the observed data given the latent variables. The objective function for an autoencoder with a latent variable layer is typically the negative log-likelihood of the observed data, which can be written as:

$$\mathcal{L}(\theta, \phi; \mathbf{X}) = -Eq\phi(\mathbf{Z}|\mathbf{X})[\log p_\theta(\mathbf{X}|\mathbf{Z})] + \mathrm{KL}(q_\phi(\mathbf{Z}|\mathbf{X}); |; p(\mathbf{Z}))$$

where $\mathbf{X}$ is the observed data, $\mathbf{Z}$ is the vector of latent variables, $p_\theta(\mathbf{X}|\mathbf{Z})$ is the likelihood function of the decoder, $q_\phi(\mathbf{Z}|\mathbf{X})$ is the variational distribution of the encoder, $p(\mathbf{Z})$ is the prior distribution over the latent variables, and θ and ϕ are the parameters of the decoder and encoder, respectively.

The first term in the objective function is the negative log-likelihood of the observed data, which measures the reconstruction error of the autoencoder. The second term is the KL divergence between the variational distribution and the prior, which encourages the learned latent variables to be close to the prior distribution. The objective function can be optimized using stochastic gradient descent or a similar optimization algorithm.

In summary, latent variable modeling and the ELBO are important tools for modeling complex data distributions with unobserved variables. Jensen's inequality is used to derive a lower bound on the log-likelihood of the observed data, which is used in variational inference to approximate the posterior distribution of the latent variables. Autoencoders use a latent variable layer to learn compressed representations of the observed data and can be trained using the ELBO as an objective function.