Marc Laugharn
MATH189 - Math of Big Data
Variational Autoencoders as a type of Expectation Maximization
March 9 2020

VAEs were introduced in http://arxiv.org/pdf/1312.6114.pdf

(Pretty much the only way I could understand this was with a lot of help from [0]., so most of the interpretation of the relationship between EM and VAE is taken wholly from that article)

In both EM and VAE we are trying to estimate the parameters of a model given data where "the model includes latent variables not specified in the data" [0]. The encoder is $P_\Psi(z|y)$. The model defines $P(z)$ and $P(y|z)$.

The algorithm fits the parameters Θ of the distribution $P_\Theta(z,x)$ to maximize the marginal probability $P_\Theta(x)$ [0]. EM algorithms alternate optimizing the latent distribution Ψ (E step) and optimizing Θ (M step) [0].

VAEs can be used where this cannot be done (efficiently) in closed form by doing gradient descent on the loss

$$\mathcal{L}(\Psi,\Theta,y) = E_{z\ P_\Psi(z|y)} \ln P_\Theta(z,y) + H(P_\Psi(z|y)) = \ln P_\Theta(y) - KL(P_\Psi(z|y), P_\Theta(z|y))$$

where H is the entropy and KL is the Kullback-Leibler divergence between Ψ and Θ. The gradient is estimated by sampling z from the encoder $P_\Psi(z|y)$ [0].

Code reproducing the results of this paper is available at https://github.com/cshenton/auto-encoding-variational-bayes

[0]: https://machinethoughts.wordpress.com/2017/10/02/vae-em/ ■