We have an i.i.d. dataset with latent variables per datapoint and would like to perform maximum likelihood (ML) or maximum a posteriori (MAP) inference on the parameters, and variational inference on the latent variables $z$ given observations $x$. We wish to find an algorithm, using a recognition model $q_{\phi}(z|x)$ to approximate the intractable true posterior $p_{\theta}(z|x)$, that works efficiently for a large dataset even when the marginal likelihood is intractable.

### Problem Setting

The log-likelihood can be expressed in terms of a regularization term plus a reconstruction term. The regularization term (KL divergence) depends on how good $q_{\phi}(z|x)$ can approximate $p_{\theta}(z|x)$. We will tune $\phi$ and $\theta$ in order to maximize the log-likelihood.

$$
\log p_{\theta}(x) = \int q_{\phi}(z|x) \log p_{\theta}(x|z) dz = \int q_{\phi}(z|x) \log p_{\theta}(x|z) q_{\phi}(z|x) p_{\theta}(z|x) dz
$$

$$
= \int q_{\phi}(z|x) \log \frac{p_{\theta}(x|z)}{q_{\phi}(z|x)} dz + \int q_{\phi}(z|x) \log \frac{p_{\theta}(z|x)}{q_{\phi}(z|x)} dz
$$

$$
= D_{KL}(q_{\phi}(z|x) \parallel p_{\theta}(z|x)) + \mathcal{L}
$$

$$
\mathcal{L} = D_{KL}(q_{\phi}(z|x) \parallel p_{\theta}(z|x)) + \mathbb{E}_{q_{\phi}(z|x)}(\log q_{\phi}(x|z))
$$

### The Variational Bound

Two practical estimators of the lower bound:

$$
\hat{\mathcal{L}}^A(\theta, \phi; x) = -\frac{1}{L} \sum_{l=1}^{L} \left( \log p_{\theta}(x, z^{l}) - \log q_{\phi}(z^{l}|x) \right)
$$

$$
\hat{\mathcal{L}}^B(\theta, \phi; x) = -D_{KL}(q_{\phi}(z|x) \parallel p_{\theta}(z)) + \frac{1}{L} \sum_{l=1}^{L} \left( \log p_{\theta}(x|z^{l}) \right)
$$

When the KL-divergence can be integrated analytically, we use $\mathcal{L}^B$ which typically generates less variance than $\mathcal{L}^A$.

Example: Variational Autoencoder

A neural network is used for the probabilistic encoder and the prior over the latent variables is Gaussian.

### Visualization of Learned Manifolds

Project high dimensional data to a 2 dimensional manifold.

### SGVB Estimator

Future Work

Extensions will feature both application-based [3] and theoretical [4] research along the following broad avenues:

- Variational autoencoders for automatic chemical design [3].
- The composition of robust features using denoising (variational) autoencoders [4].

### References


