Statistical Models and Computing Methods, Problem Set 2

November 6, 2020 Due 11/19/2020

Problem 1.

(1) Let X have the standard Laplace distribution. Use *importance sampling* based on 100000 draws from the standard normal as the proposal density to estimate $\mathbb{E}(X)$, $\mathbb{V}\text{ar}(X)$ and $\Pr(X > 2)$.

(2) Plot the logarithms of the importance weights as a histogram. Is the distribution of these log-importance-weights symmetric? Do you occasionally get extremely large importance weights? Extremely small ones? Which type of outliers is more worrisome?

(3) Compare your Monte Carlo estimates with the true values. Are there biases in the Monte Carlo estimates? How large are the Monte Carlo standard derivations? (Is there a theoretical formula for the standard deviation of your Monte Carlo estimator?)

(4) Let Y have the standard normal distribution. Use *importance sampling* based on 100000 draws from the standard Laplace as the proposal density to estimate $\mathbb{E}(Y)$, $\mathbb{V}\text{ar}(Y)$ and $\Pr(Y > 2)$. Repeat parts (2) and (3).

Problem 2.

Consider the following *Restricted Boltzmann Machine* (RBM) with energy function

$$p_\theta(v) = \frac{1}{Z_\theta} \sum_h \exp(-E(v,h)), \quad E(v,h) = -b^T v - c^T h - h^T W v$$

Here the model parameters are $\theta = \{b, c, W\}$

(1) Show that $p(v|h) = \prod_{i=1}^n p(v_i|h)$, $p(h|v) = \prod_{j=1}^d p(h_i|v)$

(2) Derive the derivatives of the log-likelihood function *w.r.t.* the model parameters θ

(3) Use the following code to load the MNIST data set.

```
from sklearn.datasets import fetch_openml

X, y = fetch_openml('mnist_784', version=1, return_X_y=True)
X = (X/255).astype('float32')
X_train, X_test = X[:60000,:], X[60000:,:]
```

Train your RBM on the training data set using contrastive divergence ($k = 1$), with Gibbs sampling for the energy induced distribution. Report the reconstruction error $\|v - \tilde{v}\|^2$ on the training data and the test data as a function of the number of iterations, where $\tilde{v}$ is the sample after $k = 1$ iteration of Gibbs sampling that starts at v. Do you have any interesting finding? Explain it.

(4) Generate samples from your trained RBM using Gibbs sampling with 10 independent chains, and each chain runs 200 iterations. Show the results in a 10×11 grid plot where columns correspond to samples at every 20 iterations.

Problem 3. Consider a logistic regression model with normal priors

$$y_i \sim \text{Bernoulli}(p_i),\ p_i = \frac{1}{1+\exp(-x_i^T\beta)},\quad i = 1, \ldots, n. \quad \beta \sim \mathcal{N}(0, \sigma_\beta^2)$$

where $\sigma_\beta = 1$. Download the data from the course website.

(1) Implement a Hamiltonian Monte Carlo sampler to collect 500 samples (with 500 discarded as burn-in), show the scatter plot. Test the following two strategies for the number of leapfrog steps L: (1) use a fixed L; (2) use a random one, say $\text{Uniform}(1, L_{\max})$. Do you find any difference? Explain it.
(2) Run HMC for 100000 iterations and discard the first 50000 samples as burn-in to form the ground truth. Implement stochastic gradient MCMC algorithms including SGLD, SGHMC and SGNHT. Show the convergence rate of different SGMCMC algorithms in terms of KL divergence to the ground truth as a function of iterations. You may want to use the ITE package `https://bitbucket.org/szzoli/ite-in-python/src/default/` to compute the KL divergence between two samples.