TU München
Fakultät für Informatik
PD Dr. Rudolph Triebel
John Chiotellis

Machine Learning for Robotics and Computer Vision
Winter term 2015

Solution Sheet 6

Topic: Sampling methods and Variational Inference
February 5th, 2016

Exercise 1: Particle Filter

a) What kind of spaces can we explore with a particle filter?

   With particle filters we can explore continuous state spaces.

b) What kind of distributions can we approximate with a particle filter?

   Particle filter is non-parametric, meaning we can approximate arbitrary distributions (Gaussian and non-Gaussian). Given enough particles we can approximate any function.

c) In a Monte Carlo localization problem what do the particles and the particle weights correspond to?

   The particles themselves correspond to the motion model as they represent the state after motion with noise. The particle weights are computed according to the measurement model so they represent the likelihood of a measurement.

d) **Programming** : Implement a particle filter for global localization.
   See code.

Exercise 2: Gibbs sampling

Show that the Gibbs sampling algorithm satisfies detailed balance:

$$p^*(z)T(z,z') = p^*(z')T(z',z)$$

This follows from the fact that in Gibbs sampling, we sample a single variable, z_k at each time, while all other variables, $\{z_i\}_{i\neq k}$, remain unchanged. Thus, $\{z'_i\}_{i\neq k} = \{z_i\}_{i\neq k}$ and we get

$$\begin{aligned}
p^*(z)T(z,z') &= p^*(z_k, \{z_i\}_{i\neq k}) p^*(z'_k | \{z_i\}_{i\neq k}) \\
&= p^*(z_k | \{z_i\}_{i\neq k}) p^*(\{z_i\}_{i\neq k}) p^*(z'_k | \{z_i\}_{i\neq k}) \\
&= p^*(z_k | \{z'_i\}_{i\neq k}) p^*(\{z'_i\}_{i\neq k}) p^*(z'_k | \{z'_i\}_{i\neq k}) \\
&= p^*(z_k | \{z'_i\}_{i\neq k}) p^*(z'_k, \{z'_i\}_{i\neq k}) \\
&= T(z',z) p^*(z'),
\end{aligned}$$

where we have used the product rule together with $T(z,z') = p^*(z'_k|\{z_i\}_{i\neq k})$.

Exercise 3: Kullback-Leibler divergence

a) What does the KL divergence describe? Is it symmetric? Why?

The Kullback-Leibler divergence is a measure of (dis)similarity between probability distributions. It is the amount of information lost when a distribution q is used to approximate a distribution p. It is minimized (zero) when the two distributions are identical. It is not symmetric. One can see that by the definition:

$$KL(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx$$

$$\begin{aligned}
KL(p||q) - KL(q||p) &= \int p(x) \log\{\frac{p(x)}{q(x)}\} dx - \int q(x) \log\{\frac{q(x)}{p(x)}\} dx \\
&= \int p(x) \log\{\frac{p(x)}{q(x)}\} dx + \int q(x) \log\{\frac{p(x)}{q(x)}\} dx \\
&= \int \log\{\frac{p(x)}{q(x)}\}(p(x) + q(x)) dx
\end{aligned}$$

which generally is non-negative, does not have to be zero.

b) Compute the KL-divergence of two univariate normal distributions.
What if they have the same mean? What if they have the same variance?

Let us define $p_1(x) = \mathcal{N}(x|\mu_1, \sigma_1)$ and $p_2(x) = \mathcal{N}(x|\mu_2, \sigma_2)$. We then have

$$KL(p_1||p_2) = \int p_1(x) \log\{\frac{p_1(x)}{p_2(x)}\} dx$$

First let us simplify the fraction

$$\begin{aligned}
\frac{p_1(x)}{p_2(x)} &= \frac{\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp(-\frac{(x-\mu_1)^2}{2\sigma_1^2})}{\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp(-\frac{(x-\mu_2)^2}{2\sigma_2^2})} = \frac{\sigma_2}{\sigma_1} \frac{\exp(-\frac{(x-\mu_1)^2}{2\sigma_1^2})}{\exp(-\frac{(x-\mu_2)^2}{2\sigma_2^2})} \\
&= \frac{\sigma_2}{\sigma_1} \exp(-\frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2})
\end{aligned}$$

Taking the logarithm of this gives us

$$\log(\frac{p_1(x)}{p_2(x)}) = \log(\frac{\sigma_2}{\sigma_1}) + \left(\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}\right)$$

Now plugging this in the KL-divergence definition we get

$$KL(p_1||p_2) = \int p_1(x)\log(\frac{\sigma_2}{\sigma_1})dx + \int p_1(x)\left(\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}\right)dx$$
$$= \log(\frac{\sigma_2}{\sigma_1})\int p_1(x)dx + \int p_1(x)\frac{(x-\mu_2)^2}{2\sigma_2^2}dx - \int p_1(x)\frac{(x-\mu_1)^2}{2\sigma_1^2}dx$$
$$= \log(\frac{\sigma_2}{\sigma_1}) + \frac{1}{2\sigma_2^2}\int p_1(x)(x-\mu_2)^2dx - \frac{1}{2\sigma_1^2}\int p_1(x)(x-\mu_1)^2dx$$
$$= \log(\frac{\sigma_2}{\sigma_1}) + \frac{1}{2\sigma_2^2}\int p_1(x)(x-\mu_1+\mu_1-\mu_2)^2dx - \frac{\sigma_1^2}{2\sigma_1^2}$$
$$= \log(\frac{\sigma_2}{\sigma_1}) + \frac{1}{2\sigma_2^2}\left(\int p_1(x)(x-\mu_1)^2dx + 2\int p_1(x)(x-\mu_1)(\mu_1-\mu_2)dx + \int p_1(x)(\mu_1-\mu_2)^2dx\right) - \frac{1}{2}$$
$$= \log(\frac{\sigma_2}{\sigma_1}) + \frac{1}{2\sigma_2^2}\left(\sigma_1^2 + 2(\mu_1-\mu_2)\int p_1(x)(x-\mu_1)dx + (\mu_1-\mu_2)^2\int p_1(x)dx\right) - \frac{1}{2}$$
$$= \log(\frac{\sigma_2}{\sigma_1}) + \frac{1}{2\sigma_2^2}\left(\sigma_1^2 + (\mu_1-\mu_2)^2\right) - \frac{1}{2}$$

If two distributions only differ in their mean values ($\sigma_1 = \sigma_2$) then the KL-divergence is proportional to the square of their means difference,

$$KL(p||q) = \frac{(\mu_1-\mu_2)^2}{2\sigma_2^2}.$$

If they have equal mean but different variances ($\mu_1 = \mu_2$) then the KL-divergence is a function of the ratio of their variances:

$$KL(p||q) = \log(\frac{\sigma_2}{\sigma_1}) + \frac{\sigma_1^2}{2\sigma_2^2} - \frac{1}{2} = \frac{\sigma_1^2}{2\sigma_2^2} - \log(\frac{\sigma_1}{\sigma_2}) - \frac{1}{2}$$

c) Consider a factorized variational distribution $q(Z)$. By using the technique of Lagrange multipliers, verify that minimization of $KL(p||q)$ with respect to one of the factors $q_i(Z_i)$ keeping all other factors fixed, leads to the solution:

$$q_j^*(Z_j) = \int p(Z)\prod_{i\neq j} dZ_i = p(Z_j)$$

$$
\begin{aligned}
KL(p||q) &= \int p(Z) \ln \frac{p(Z)}{q(Z)} dZ \\
&= \int p(Z) \ln p(Z) dZ - \int p(Z) \ln q(Z) dZ \\
&= \int p(Z) \ln p(Z) dZ - \int p(Z) \ln \prod_i q_i(Z_i) dZ \\
&= -\int p(Z) \sum_{i=1}^{M} \ln q_i(Z_i) dZ + const. \\
&= -\int (p(Z) \ln q_j(Z_j) + p(Z) \sum_{i \neq j} \ln q_i(Z_i)) dZ + const. \\
&= -\int p(Z) \ln q_j(Z_j) dZ + const. \\
&= -\int \ln q_j(Z_j) \left(\int p(Z) \prod_{i \neq j} dZ_i \right) dZ_j + const.
\end{aligned}
$$

We want to minimize this and at the same time enforce the constraint

$$\int q_j(Z_j) dZ_j = 1.$$

Therefore we add a Lagrange multiplier and our objective function becomes

$$f(q_j(Z_j)) = -\int \ln q_j(Z_j) \left(\int p(Z) \prod_{i \neq j} dZ_i \right) dZ_j + \lambda \left(\int q_j(Z_j) dZ_j - 1 \right)$$

Taking the derivative w.r.t. $q_j(Z_j)$ and setting it equal to zero we get

$$\frac{\partial f(q_j(Z_j))}{\partial q_j(Z_j)} = -\frac{\int p(Z) \prod_{i \neq j} dZ_i}{q_j(Z_j)} + \lambda \overset{!}{=} 0$$

We solve for λ

$$
\begin{aligned}
\lambda q_j(Z_j) &= \int p(Z) \prod_{i \neq j} dZ_i \\
\lambda \int q_j(Z_j) dZ_j &= \int \left(\int p(Z) \prod_{i \neq j} dZ_i \right) dZ_j \\
\lambda &= 1
\end{aligned}
$$

And thus

$$q_j(Z_j) = \int p(Z) \prod_{i \neq j} dZ_i$$