Importance Sampling

Alena Shilova

Skoltech

June 05,2017

Normal distribution

For the normal distribution $\mathcal{N}(\mu, \Sigma)$ to look like the one from exponential family, let's make a change of variables: $S = \Sigma^{-1}$ and $m = \Sigma^{-1}\mu$. Then the density of distribution will look like the following:

$$g_{m,S}(x) = \frac{1}{2\pi} \exp\Big(m^T x - \frac{1}{2}\operatorname{tr}(Sxx^T)\Big) \exp\Big(-\frac{1}{2}\big(m^T S^{-1} m - \log|S|\big)\Big).$$

So in this case $A(m, S) = \frac{1}{2} m^T S^{-1} m - \frac{1}{2}\log|S|$, $T(x) = (x, -\frac{1}{2}xx^T)$ and $h(x) = \frac{1}{2\pi}$

Strong convexity

It is a well-known fact that if $f(x)$ is strong convex with a coefficient β, then $e^{f(x)}$ is a strongly convex function on some compact and if $x \in [-C, C]$ for some $C > 0$, then the coefficient of strong convexity is equal to βe^{-C}. In our case if $\lambda(\Sigma) \in [\lambda_{min}, \lambda_{max}]$, then

$$\frac{1}{2}x^T S x - m^T x + \frac{1}{2}m^T S^{-1} m - \frac{1}{2}\log|S| \geq -\frac{1}{2}\log|S| \geq \frac{n}{2}\log\lambda_{min}$$

Therefore, $\alpha = \lambda_{min}^{\frac{n}{2}}$. Now we need to find β.

Strong convexity

It is quite obvious that $\frac{1}{2}x^T Sx - m^T x$ is convex w.r.t. distribution parameters m, S. One can easily check the convexity of $f(x) = \frac{1}{2}m^T S^{-1} m$ using the following criterion:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$$

Indeed, here $\nabla f(x) = (S^{-1}m, -\frac{1}{2}S^{-1}mm^T S^{-1})$, so

$$\frac{1}{2}\operatorname{tr}\left(\left(S_1^{-1}m_1 m_1^T S_1^{-1} - S_2^{-1}m_2 m_2^T S_2^{-1}\right)^T \left(S_2 - S_1\right)\right) +$$

$$+\left(S_1^{-1}m_1 - S_2^{-1}m_2\right)^T \left(m_1 - m_2\right) = \frac{1}{2}(x_1 - x_2)^T (S_1 + S_2)(x_1 - x_2) > 0$$

where $x_i = S_i^{-1}m_i$ and last inequality holds due to S_i is positive-definite for any $i \in \{1, 2\}$

Strong convexity

Let's prove the strong convexity of the function $f(S) = -\frac{1}{2}\log|S|$.

Theorem (about strong/smooth duality)

$f(S)$ is β-strongly convex w.r.t. a norm $\|\cdot\|$ if and only if f^* is $\frac{1}{\beta}$-strongly smooth w.r.t. the dual norm $\|\cdot\|_*$

The dual function is $f^*(S) = -2n + \frac{1}{2}\ln 2 - \frac{1}{2}\sum_{i=1}^{n}\ln|\lambda_i(S)|$, where S is a negative-definite matrix with eigenvalues lying in the interval $\left[-\frac{\lambda_{max}}{2}, -\frac{\lambda_{min}}{2}\right]$, i.e. has only negative eigenvalues. λ_{max} and λ_{min} are respectively maximal and minimal eigenvalues of the covariance matrix of the considered normal distribution.
To prove its strong smoothness, I'm going to use lemma of Juditsky and Nemirovski (2008).

Juditsky and Nemirovski

Lemma

Let δ be an open interval. Suppose $\phi : \delta \to \mathbb{R}_*$ is a twice differentiable convex function such that ϕ'' is monotonically non-decreasing. Let $\mathbb{S}_n(\delta)$ be the set of all symmetric $n \times n$ matrices with eigenvalues in δ. Define the function $F : \mathbb{S}_n(\delta)\dagger'\mathbb{R}_*$

$$F(X) = \sum_{i=1}^{n} \phi(\lambda_i(X))$$

and let

$$f(t) = F(X + tH)$$

for some $X \in \mathbb{S}_n(\delta)$, $H \in \mathbb{S}_n$. Then, we have,

$$f''(0) \leq 2 \sum_{i=1}^{n} \phi''(\lambda_i(X))\lambda_i^2(H)$$

Strong convexity

Let's use this Lemma and consider
$f(t) = \frac{1}{2}\ln 2 - 2n - \frac{1}{2}\sum_{i=1}^{n} \ln|\lambda_i(X + tH)|$ for some fixed X, H.
Applying lemma on it:

$$f''(0) \leq 2\sum_{i=1}^{n} \frac{\lambda_i^2(H)}{2\lambda_i^2(X)} \leq \frac{4}{\lambda_{min}^2}\sum_{i=1}^{n} \lambda_i^2(H) = \frac{4}{\lambda_{min}^2}\|H\|_F^2$$

It means that f(S) is $\frac{\lambda_{min}^2}{4}$ −strong convex.

Finally, $g(m, S)$ is $\frac{\lambda_{min}^{\frac{n}{2}+2}}{4}$ −strong convex

Closing remark

If we have a proper $\alpha(x)$ (here $\alpha(x) = \frac{1}{4}\lambda_{min}^{\frac{n}{2}+2}$) and a set $\mathbf{\Theta}$, then for all $\theta_1, \theta_2 \in \mathbf{\Theta}$ we will have:

$$\int \exp(A(t(\theta_1) + (1-t)\theta_2) - (t\theta_1 + (1-t)\theta_2)^T T(x)) f^2(x)\phi^2(x)h(x)\,dx \leq$$
$$\leq t \int \exp(A(\theta_1) - \theta_1^T T(x)) f^2(x)\phi^2(x)h(x)\,dx +$$
$$+(1-t) \int \exp(A(\theta_2) - \theta_2^T T(x)) f^2(x)\phi^2(x)h(x)\,dx -$$
$$-t(1-t)\frac{\lambda_{min}^{\frac{n}{2}+2}}{8\pi} \int f^2(x)\phi^2(x)\,dx \|\theta_1 - \theta_2\|_2^2$$

The coefficient will depend on the value of $\int\limits_{\mathbb{R}^n} f^2(x)\phi^2(x)\,dx$, but we can also take $\alpha(x) = \frac{1}{4}\lambda_{min}^{\frac{n}{2}+2}\mathbb{I}(x \in X)$ for some X and then it is enough to know the value $\int\limits_X f^2(x)\phi^2(x)\,dx$