STA 4273 Topics in Statistical Learning Theory

Assignment - 1

Hitarth Choubisa (1004965479)

December 26, 2019

1. Gaussian Mean Estimation

1.1

Estimator $\hat{\mu}^s$ is defined as $\hat{\mu}^s = (1 - \frac{\tau}{||\hat{\mu}||_2^2})\hat{\mu}$, where $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$. We want to find the optimal τ that minimizes the risk $R(\hat{\mu}^s, \hat{\mu})$.

$$R(\hat{\mu}^s, \hat{\mu}) = E[||\hat{\mu}^s - \mu||_2^2] = E[||\hat{\mu} - \mu + g(\hat{\mu})||_2^2]$$

where $g(x) = \frac{-\tau}{||x||^2}x$. Expanding the above expression we get,

$$R(\hat{\mu}^s, \hat{\mu}) = E[||\hat{\mu} - \mu||_2^2] + E[||g(\hat{\mu})||_2^2] + 2 \cdot E[\langle \hat{\mu} - \mu, g(\hat{\mu}) \rangle]$$

$$= \frac{\sigma^2 d}{n} + E[\frac{\tau^2}{||x||_2^2}] + \frac{\sigma^2 d}{n} E[Tr(\nabla g(\hat{\mu}))]$$

where we used Stein's Lemma in simplifying the last term. Here,

$$\nabla g(x) = \frac{-\tau}{||x||^2} I + \frac{2\tau x x^T}{||x||^4} \implies Tr(\nabla g(x)) = -\frac{\tau(d-2)}{||x||_2^2}$$

Thus, we get expression of risk as,

$$R(\hat{\mu}^s, \hat{\mu}) = \frac{\sigma^2 d}{n} + \tau^2 E[\frac{1}{||x||_2^2}] - \tau \frac{\sigma^2 d(d-2)}{n} E[\frac{1}{||x||_2^2}]$$

Value for which the above expression is minimized can be obtain by setting its derivative wrt τ as zero,

$$2\tau E[\frac{1}{||x_2^2||}] = \frac{\sigma^2 d(d-2)}{n} E[\frac{1}{||x_2^2||}] \implies \tau_{opt} = \frac{\sigma^2 d(d-2)}{2n}$$

1.2(a)

$$E[\nabla_x log(p_\eta(X)) g_\eta(X)^T] + E[\nabla_x g_\eta(X)] = \int \frac{d}{dx} log(p_\eta(x) \cdot g_\eta(x)^T \cdot p_\eta(x) dx + E[\nabla_x g_\eta(X)]$$

$$= \int \frac{1}{p_\eta(x)} \frac{d(p_\eta(x))}{dx} \cdot g_\eta(x)^T \cdot p_\eta(x) dx + E[\nabla_x g_\eta(X)]$$

$$= \int \frac{d(p_\eta(x))}{dx} g_\eta(x)^T dx + E[\nabla_x g_\eta(X)] = g_\eta^T(x) p_\eta(x)|_{-\infty}^{\infty} - \int \nabla g(x) p_\eta(x) dx + E[\nabla_x g_\eta(X)]$$

Since $g_\eta(x)$ and $p_\eta(x)$ are differential functions, both of them approach zero as $x \longrightarrow 0$. Thus, the above expression evaluates to,

$$= -E[\nabla_x g_\eta(x)] + E[\nabla_x g_\eta(X)] = 0$$

1.2(b)

$$E[\nabla_\eta log(p_\eta(X))g_\eta(X)] + E[\nabla_\eta g_\eta(X)] = \int \frac{d(log(p_\eta(x)))}{d\eta} \cdot g_\eta(x)^T \cdot p_\eta(x)dx + \int \frac{d}{d\eta}(g_\eta(x)) \cdot p_\eta(x)dx$$
$$= \int \frac{1}{p_\eta(x)} \cdot \frac{dp_\eta(x)}{d\eta} \cdot g_\eta(x)^T p_\eta(x)dx + \int \frac{d}{d\eta}(g_\eta(x)) \cdot p_\eta(x)dx$$
$$= \int \frac{d}{d\eta}(p_\eta(x))g_\eta(x)^T dx + \int \frac{d}{d\eta}(g_\eta(x)) \cdot p_\eta(x)dx$$
$$= \int \frac{d}{d\eta}(p_\eta(x)g_\eta(x))dx = \frac{d}{d\eta}\int g_\eta(x)p_\eta(x)dx = \frac{d}{d\eta}\xi(\eta) = \nabla_\eta\xi(\eta)$$

1.3

$$E[||\hat{\mu}(x) - \mu||_2^2] = E[||x - \mu + g(x)||_2^2] = E[||x - \mu||_2^2] + E[||g(x)||_2^2] + 2 \cdot E[\langle x - \mu, g(x)\rangle]$$
$$= E[Tr[(x-\mu)(x-\mu)^T]] + E[||g(x)||_2^2] + 2 \cdot E[\langle x - \mu, g(x)\rangle]$$

Using results from 1.2(a) we get,

$$E[\nabla_x log(p_\eta(X))g_\eta(X)^T] = -E[\nabla_x g_\eta(X)]$$
$$\implies E[\nabla_x(\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))g_\eta(X)^T] = -E[\nabla_x g_\eta(X)]$$
$$\implies E[(x-\mu)\Sigma^{-1}g_\eta(X)^T] = E[(x-\mu)g_\eta(X)^T\Sigma^{-1}] = E[\nabla_x g_\eta(X)]$$
$$\implies E[(x-\mu)g_\eta(X)^T] = E[\langle x-\mu, g(x)\rangle] = E[Tr(\Sigma\nabla_x g_\eta(X))]$$

Plugging the above value in the expression derived earlier, we get,

$$E[||\hat{\mu}(x) - \mu||_2^2] = Tr[E[(x-\mu)(x-\mu)^T]] + E[||g(x)||_2^2] + 2 \cdot E[Tr(\Sigma\nabla_x g_\eta(X))]$$
$$= E[Tr(\Sigma)] + E[||g(x)||_2^2] + 2 \cdot E[Tr(\Sigma\nabla_x g_\eta(X))]$$
$$= E[Tr(\Sigma) + ||g(x)||_2^2 + 2 \cdot Tr(\Sigma\nabla_x g_\eta(X))] = E[S(X, \hat{\mu})]$$

2. Exponential families

2.1

$$Tr[E[(\phi(x)-\xi)(\phi(x)-\xi)^T]] = Tr(E[\phi(x)\phi(x)^T] - E[\phi(x)\xi^T] - E[\xi\phi(x)^T] + E[\xi\xi^T])$$
$$= Tr(E[\phi(x)]E[\phi(x)]^T + Cov(\phi(X)) - 2 \cdot E[\phi(x)]\xi^T + E[\xi\xi^T])$$
$$= Tr(\nabla_\eta\psi(\eta)\nabla_\eta\psi(\eta)^T + \nabla^2\psi(\eta) - 2\nabla\psi(\eta)\xi^T + \xi\xi^T)$$

since $E[\phi(X)] = \nabla\psi(\eta)$ and $Cov(\phi(X)) = \nabla^2\psi(\eta)$.

2.2(a)

$$E[\nabla_\eta l_\eta(X)] = \int \frac{d}{d\eta} log(p_\eta(x)) \cdot p_\eta(x)dx$$
$$= \int \frac{1}{p_\eta(x)} p_\eta(x) \frac{d}{d\eta} p_\eta(x)dx = \frac{d}{d\eta}\int p_\eta(x)dx = \frac{d}{d\eta} \cdot 1 = 0$$

2.2(b)

Using result from 1.2 wherein $g_\eta(X) = \nabla_\eta l_\eta(X)$ we get,

$$E[\nabla_\eta l_\eta(X)\nabla_\eta l_\eta(X)^T] = -E[\nabla_\eta g_\eta(X)]$$

since $E[g_\eta(X)] = E[\nabla_\eta l_\eta(X)] = 0$ from the previous subpart which implies,

$$= -E[\nabla_\eta^2 l_\eta(X)]$$