Machine Learning (Homework #1 Solution)

Problem 1

$p(t|x,\mathbf{x},\mathbf{t}) = \int_{-\infty}^{\infty} p(t|x,\mathbf{w},\beta)p(\mathbf{w}|\mathbf{x},\mathbf{t})d\mathbf{w}$

First step

$p(\mathbf{w}|\mathbf{x},\mathbf{t}) \propto p(\mathbf{t}|\mathbf{x},\mathbf{w})p(\mathbf{w}|\alpha)$

by equations in page 93

$p(\mathbf{t}|\mathbf{x},\mathbf{w}) = \mathcal{N}(\mathbf{t}|\mathbf{w}^T\mathbf{\Phi}(x),\beta^{-1}\mathbf{I}) = \mathcal{N}(\mathbf{t}|\mathbf{w}^T\mathbf{A}+b,\mathbf{L}^{-1})$

$\rightarrow \mathbf{A} = \mathbf{\Phi}(x)^T, b = 0, \mathbf{L} = \beta\mathbf{I}$

$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|0,\alpha^{-1}\mathbf{I}) = \mathcal{N}(\mathbf{w}|\mu,\mathbf{\Lambda}^{-1}) \rightarrow \mu = 0, \mathbf{\Lambda} = \alpha\mathbf{I}$

$p(\mathbf{w}|\mathbf{x},\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{\Sigma}\{\mathbf{A}^T\mathbf{L}(\mathbf{w}-b)+\mathbf{\Lambda}\mu\},\mathbf{\Sigma})$, where $\mathbf{\Sigma} = (\alpha\mathbf{I}+\mathbf{A}^T\mathbf{L}\mathbf{A})^{-1}$

substitute $\mathbf{A} = \mathbf{\Phi}(x)^T, b = 0, \mathbf{L} = \beta\mathbf{I}, \mu = 0, \Lambda = \alpha\mathbf{I}$

$\rightarrow \mathcal{N}(\mathbf{w}|\mathbf{S}(\mathbf{\Phi}^T(x)\beta\mathbf{t}),\mathbf{S})$, where $\mathbf{S} = (\alpha\mathbf{I}+\mathbf{\Phi}(x)\beta\mathbf{\Phi}(x)^T)^{-1}$

Second step

by equations in page 93

$p(t|\mathbf{w},\mathbf{x}) = \mathcal{N}(t|\mathbf{w}^T\mathbf{\Phi}(x),\beta^{-1}) = \mathcal{N}(t|\mathbf{w}^T\mathbf{A}+b,\mathbf{L}^{-1}) \rightarrow \mathbf{A} = \mathbf{\Phi}(x), b = 0, \mathbf{L} = \beta\mathbf{I}$

$p(\mathbf{w}|\mathbf{x},\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{S}(\beta\mathbf{\Phi}(x)\mathbf{t}),\mathbf{S}) = p(\mathbf{w}|\mu,\mathbf{\Lambda}^{-1}) \rightarrow \mu = \mathbf{S}(\beta\mathbf{\Phi}(x)\mathbf{t}), \mathbf{\Lambda}^{-1} = \mathbf{S}$

substitute $\mathbf{A} = \mathbf{\Phi}(x)^T, b = 0, \mathbf{L} = \beta\mathbf{I}, \mu = 0, \Lambda = \alpha\mathbf{I}$

$p(t|x,\mathbf{x},\mathbf{t}) = \mathcal{N}(t|\mathbf{A}\boldsymbol{\mu}+b,\mathbf{L}^{-1}+\mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^T)$

$= \mathcal{N}(t|\beta\mathbf{\Phi}(x)^T\mathbf{S}\mathbf{\Phi}(x)\mathbf{t},\beta^{-1}+\mathbf{\Phi}(x)^T\mathbf{S}\mathbf{\Phi}(x))$

Problem 2.

method 1

step 0

To show that maximum entropy distribution for a continuous variable with three constraints

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

$$\int_{-\infty}^{\infty} xp(x)dx = \mu$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx = \sigma^2$$

is a Gaussian distribution. We need to know the following facts first

✓ $\int_{-\infty}^{\infty} e^{-a(x+b)^2}dx = \sqrt{\frac{\pi}{a}}$ (first fact)

✓ $\int_{-\infty}^{\infty} xe^{-a(x-b)^2}dx = b\sqrt{\frac{\pi}{a}}$ (second fact)

✓ $\int_{-\infty}^{\infty} x^2e^{-ax^2}dx = \frac{1}{2}\sqrt{\frac{\pi}{a^3}}$ (third fact)

step 1

We want to maximize the following functional with respect to $p(x)$

$$-\int_{-\infty}^{\infty} p(x)\log p(x)dx + \lambda_1\Big(\int_{-\infty}^{\infty} p(x)dx - 1\Big) + \lambda_2\Big(\int_{-\infty}^{\infty} xp(x)dx - \mu\Big) + \lambda_3\Big(\int_{-\infty}^{\infty} (x-\mu)^2p(x)dx - \sigma$$

So we take derivative directly w.r.t $p(x)$ and set to zero, which yields

$$\begin{aligned}
&-(\log p(x)+1) + \lambda_1 + \lambda_2 x + \lambda_3(x-\mu)^2 = 0\\
&\Rightarrow p(x) = \exp(-1+\lambda_1+\lambda_2 x+\lambda_3(x-\mu)^2)\\
&\Rightarrow p(x) = \exp(\lambda_3\big[x+\frac{\lambda_2-2\mu\lambda_3}{2\lambda_3}\big]^2)\times\exp(\lambda_1-1-\frac{\lambda_2^2}{4\lambda_3}+\mu\lambda_2)
\end{aligned}$$

step 2

Using the first constraints and the first fact, we can obtain

$$\begin{aligned}
&\int_{-\infty}^{\infty}\exp(\lambda_3\big[x+\frac{\lambda_2-2\mu\lambda_3}{2\lambda_3}\big]^2)dx\times\exp(\lambda_1-1-\frac{\lambda_2^2}{4\lambda_3}+\mu\lambda_2) = 1\\
&\Rightarrow \sqrt{\frac{\pi}{-\lambda_3}}\times\exp(\lambda_1-1-\frac{\lambda_2^2}{4\lambda_3}+\mu\lambda_2) = 1\\
&\Rightarrow \exp(\lambda_1-1-\frac{\lambda_2^2}{4\lambda_3}+\mu\lambda_2) = \sqrt{\frac{-\lambda_3}{\pi}}
\end{aligned}$$

step 3

Using the second constraints and the second fact, we can obtain

$$\int_{-\infty}^{\infty} x \exp(\lambda_3 [x + \frac{\lambda_2 - 2\mu\lambda_3}{2\lambda_3}]^2) dx \times \sqrt{\frac{-\lambda_3}{\pi}} = \mu$$
$$\Rightarrow \frac{\lambda_2 - 2\mu\lambda_3}{-2\lambda_3} \sqrt{\frac{\pi}{-\lambda_3}} \sqrt{\frac{-\lambda_3}{\pi}} = \mu$$
$$\Rightarrow \frac{\lambda_2}{-2\lambda_3} + \mu = \mu$$
$$\Rightarrow \lambda_2 = 0$$

step 4

Using the third constraints, the third fact, $\lambda_2 = 0$ and let $u = x - \mu$, we can obtain

$$\int_{-\infty}^{\infty} (x-\mu)^2 \exp(\lambda_3 [x + \frac{\lambda_2 - 2\mu\lambda_3}{2\lambda_3}]^2) dx \times \sqrt{\frac{-\lambda_3}{\pi}} = \sigma^2$$
$$\Rightarrow \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left[\lambda_3 (x-\mu)^2\right] dx \times \sqrt{\frac{-\lambda_3}{\pi}} = \sigma^2$$
$$\Rightarrow \int_{-\infty}^{\infty} u^2 \exp(\lambda_3 u^2) du \times \sqrt{\frac{-\lambda_3}{\pi}} = \sigma^2$$
$$\Rightarrow \frac{1}{2} \sqrt{\frac{\pi}{-\lambda_3^3}} \sqrt{\frac{-\lambda_3}{\pi}} = \sigma^2$$
$$\Rightarrow \lambda_3 = \frac{-1}{2\sigma^2}$$

step 5

$$\exp(\lambda_1 - 1) = \sqrt{\frac{-\lambda_3}{\pi}} \Rightarrow \exp(\lambda_1 - 1) = \sqrt{\frac{1}{2\pi\sigma^2}}$$
$$\Rightarrow \lambda_1 - 1 = -\log 2\pi\sigma^2$$
$$\Rightarrow \lambda_1 = 1 - \log 2\pi\sigma^2$$

method 2

step 1

Assume that $p(x)$ is a normal distribution. We know that KL-divergence is always larger than or equal to zero

$$0 \leq \mathbb{KL}(q||p) = \int q(x) \log \frac{q(x)}{p(x)} dx$$
$$= -h(q) - \int q(x) \log p(x) dx$$
$$\Rightarrow h(q) \leq - \int q(x) \log p(x) dx$$

step 2

Note that

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

$$\int q(x) \log p(x) dx = - \int_{-\infty}^{\infty} q(x) \left[\frac{1}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2}(x-\mu)^2\right] dx$$
$$= -\frac{1}{2} \log(2\pi\sigma^2) \int_{-\infty}^{\infty} q(x) dx - \frac{1}{2\sigma^2} \int_{-\infty}^{\infty} q(x)(x-\mu)^2 dx$$
$$= -\frac{1}{2} \log(2\pi\sigma^2) \int_{-\infty}^{\infty} p(x) dx - \frac{1}{2\sigma^2} \int_{-\infty}^{\infty} p(x)(x-\mu)^2 dx$$
$$= -\frac{1}{2} \log(2\pi\sigma^2) \int_{-\infty}^{\infty} p(x) dx - \frac{1}{2\sigma^2} \int_{-\infty}^{\infty} p(x)(x-\mu)^2 dx$$
$$= - \int_{-\infty}^{\infty} p(x) \left[\frac{1}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2}(x-\mu)^2\right] dx$$
$$= \int p(x) \log p(x) dx$$

step 3

Combine the result

$$h(q) \leq - \int q(x) \log p(x) dx = - \int p(x) \log p(x) dx = h(p)$$