Machine Learning 1: Assignment 4

Tom Nick 340528
Niklas Gebauer 340942

Exercise 1

(a) The expected value of a continuous random variable x with density function $p(x)$ is given by:
$E[x] = \int x \cdot p(x)dx$
The variance is given by:
$Var[x] = E[x^2] - (E[x])^2$
In our case the expected value (which is also the mean) is supposed to be zero, thus we can rewrite the variance as:
$Var[x] = E[x^2] = \int x^2 \cdot p(x)dx$

Using those equations we get the following Lagrangian:

$$\begin{aligned}\Lambda(s(x), \lambda_1, \lambda_2, \lambda_3) &= -\int e^{s(x)}s(x)dx + \lambda_1(\int e^{s(x)}dx - 1) + \lambda_2 \int e^{s(x)}xdx + \lambda_3(\int e^{s(x)}x^2dx - \sigma^2) \\ &= \int -e^{s(x)}s(x) + \lambda_1 e^{s(x)} + \lambda_2 e^{s(x)}x + \lambda_3 e^{s(x)}x^2dx - \lambda_1 - \lambda_3\sigma^2\end{aligned}$$

(b) We take the gradient of our Lagrangian to show that the function $s(x)$ that maximizes the objective $H(x)$ is quadratic in x.

$$\nabla\Lambda = \begin{pmatrix} \int -e^{s(x)}(1 + s(x)) + e^{s(x)}\lambda_1 + e^{s(x)}\lambda_2 x + e^{s(x)}\lambda_3 x^2dx \\ \int e^{s(x)}dx - 1 \\ \int e^{s(x)}xdx \\ \int e^{s(x)}x^2dx \end{pmatrix} \stackrel{!}{=} \vec{0}$$

If we analyze the first component we can show that $s(x)$ will be quadratic in x:

$$\begin{aligned}\int -e^{s(x)}(1 + s(x)) + e^{s(x)}\lambda_1 + e^{s(x)}\lambda_2 x + e^{s(x)}\lambda_3 x^2dx &= 0 \\ \Leftrightarrow \int e^{s(x)}(\lambda_3 x^2 + \lambda_2 x + \lambda_1 - 1 - s(x))dx &= 0 \\ \Leftrightarrow \int e^{s(x)}(\lambda_3 x^2 + \lambda_2 x + \lambda_1 - 1)dx &= \int e^{s(x)}s(x)dx \\ \Leftrightarrow \lambda_3 x^2 + \lambda_2 x + \lambda_1 - 1 &= s(x)\end{aligned}$$

(c) Using the formula from above and $p(x) = e^{s(x)}$ we can obtain:

$$p(x) = e^{\lambda_3 x^2 + \lambda_2 x + \lambda_1 - 1}$$

We will now use the constraints to show that the function $p(x)$ that maximizes $H(x)$ is a Gaussian distribution with mean $\mu = 0$ and variance σ^2. Therefore we use the following to equations for Gaussian integrals:

$$\int e^{-ax^2+bx+c}dx = \sqrt{\frac{\pi}{a}}e^{\frac{b^2}{4a}+c} \tag{1}$$

$$\int x^{2n}e^{-ax^2+bx+c}dx = \frac{(2n-1)!!}{(2a)^n}\sqrt{\frac{\pi}{a}}e^{\frac{b^2}{4a}+c} \tag{2}$$

Using the first constraint we get:

$$\int e^{s(x)}dx - 1 = 0$$
$$\Leftrightarrow \int e^{\lambda_3 x^2 + \lambda_2 x + \lambda_1 - 1}dx = 1$$
$$(1) \Leftrightarrow \sqrt{\frac{\pi}{-\lambda_3}} e^{\frac{\lambda_2^2}{-4\lambda_3} + (\lambda_1 - 1)} = 1$$

Using this and the third constrain we get:

$$\int e^{s(x)} x^2 dx = \sigma^2$$
$$(2) \Leftrightarrow \frac{1!!}{-2\lambda_3} \sqrt{\frac{\pi}{-\lambda_3}} e^{\frac{\lambda_2^2}{-4\lambda_3} + (\lambda_1 - 1)} = \sigma^2$$
$$\Leftrightarrow -\frac{1}{2\lambda_3} = \sigma^2$$
$$\Leftrightarrow \lambda_3 = -\frac{1}{2\sigma^2}$$

We can plug this into the first constraint and obtain:

$$\sqrt{\frac{\pi}{-\lambda_3}} e^{\frac{\lambda_2^2}{-4\lambda_3} + (\lambda_1 - 1)} = 1$$
$$\Leftrightarrow \sqrt{2\pi\sigma^2} e^{\lambda_2 2\sigma^2 + \lambda_1 - 1} = 1$$
$$\Leftrightarrow \lambda_1 - 1 = -ln(\sqrt{2\pi\sigma^2}) - \lambda_2^2\sigma^2$$

If we plug this into our formula for p(x) we get:

$$p(x) = e^{-\frac{x^2}{2\sigma^2} + \lambda_2(x - 2\sigma^2) - ln(\sqrt{2\pi\sigma^2})}$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2} + \lambda_2(x - 2\sigma^2)}$$

Since we know that our distribution has mean zero (constraint 2) λ_2 has to be zero. Otherwise the distribution will shift away. so we finally get:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

So we have $p(x) \sim \mathcal{N}(0, \sigma^2)$.

(d) We have to show:

$$J(x) \geq 0$$
$$\Leftrightarrow H(x*) - H(x) \geq 0$$
$$\Leftrightarrow H(x*) \geq H(x)$$

We know $x* \sim \mathcal{N}(0, \sigma^2)$. From c) we know that the differential entropy for random variables with mean $\mu = 0$ (constraint 2) and variance σ^2 (constraint 3) is maximized by the Gaussian probability density function. So if $x* \sim \mathcal{N}(0, \sigma^2)$ there cannot be any other pdf of x so that $H(x) > H(x*)$.

This means that $H(x*) \geq H(x)$ holds and thus $J(x) \geq 0$. Furthermore both will be equal, if $x \sim \mathcal{N}(0, \sigma^2)$, say Gaussian distributed, since then $H(x) = H(x*)$. So $J(x) = 0$ in the case that x is Gaussian distributed.

This means that $J(x)$ will get bigger the less similar to Gaussian x is distributed.

Exercise 2