Homework #1

Shao Hua, Huang
ECM9032 - Machine Learning

October 21, 2017

1 Jensen's inequality (10%)

Proof. Use mathematical induction to prove Jensen's inequality.

(a) for $M = 2$, the inequality (1.1) holds

$$f(\sum_{i=1}^{M} \lambda_i x_i) \leq \sum_{i=1}^{M} \lambda_i f(x_i) \tag{1.1}$$

which derived from (1) in problem description

(b) assume $M = n$ holds, *i.e.*

$$f(\sum_{i=1}^{n} \lambda_i x_i) \leq \sum_{i=1}^{n} \lambda_i f(x_i) \text{ where } \lambda_i > 0 \text{ and } \sum_{i=1}^{n} \lambda_i = 1 \tag{1.2}$$

(c) for $M = n + 1$,

$$\begin{aligned}
f(\sum_{i=1}^{n+1} \lambda_i x_i) &= f(\sum_{i=1}^{n} \lambda_i x_i + \lambda_{n+1} x_{n+1}) \\
&= f\left[(1 - \lambda_{n+1}) \sum_{i=1}^{n} \frac{\lambda_i x_i}{1 - \lambda_{n+1}} + \lambda_{n+1} x_{n+1}\right] && (1.3) \\
&\leq (1 - \lambda_{n+1}) f(\sum_{i=1}^{n} \frac{\lambda_i x_i}{1 - \lambda_{n+1}}) + \lambda_{n+1} f(x_{n+1}) && (1.4) \\
&\leq (1 - \lambda_{n+1}) \sum_{i=1}^{n} \frac{\lambda_i}{1 - \lambda_{n+1}} f(x_i) + \lambda_{n+1} f(x_{n+1}) && (1.5) \\
&= \sum_{i=1}^{n} \lambda_i f(x_i) + \lambda_{n+1} f(x_{n+1}) \\
&= \sum_{i=1}^{n+1} \lambda_i f(x_i)
\end{aligned}$$

Use (1.1) to infer inequality from (1.3) to (1.4)
Besides, because of

$$\frac{\lambda_i}{1-\lambda_{n+1}} > 0 \text{ and } \sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} = 1$$

We can infer (1.4) to (1.5) with (1.2)

(d) By mathematical induction and (a), (b), and (c),
we can prove that

$$f(\sum_{i=1}^{M} \lambda_i x_i) \leq \sum_{i=1}^{M} \lambda_i f(x_i) \quad \forall M \in \mathbb{N} \wedge M > 2$$

□

2 Entropy of the univariate Gaaussian (15%)

Proof.

$$\begin{aligned}
H[x] &= \mathbb{E}[-\ln(p(x))] \\
&= \mathbb{E}\left[-\ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\exp^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)\right] && (2.1) \\
&= \mathbb{E}\left[\frac{1}{2}\ln(2\pi\sigma^2) + \frac{(x-\mu)^2}{2\sigma^2}\right] && (2.2) \\
&= \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\mathbb{E}\left[(x-\mu)^2\right] && (2.3) \\
&= \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\int_{-\infty}^{\infty}(x-\mu)^2 p(x)\mathrm{d}x && (2.4) \\
&= \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sigma^2 && (2.5) \\
&= \frac{1}{2}\left(1+\ln(2\pi\sigma^2)\right)
\end{aligned}$$

From (2.1) to (2.2), we perform basic log operation.
From (2.2) to (2.3), we take out independent terms in expectations.
To get (2.5), we substitude equation about variance in problem to (2.3).
Therefore, we prove the equation

$$H[x] = \frac{1}{2}\left(1+\ln(2\pi\sigma^2)\right)$$

□

3 KL divergence between two Gaussians (15%)

Assume both σ and s are positive.

$$\begin{aligned}
\mathrm{KL}(p \parallel q) &= -\int p(x) \ln \left(\frac{q(x)}{p(x)} \right) \mathrm{d}x \\
&= -\int p(x) \ln \left(\frac{\frac{1}{\sqrt{2\pi s^2}} \exp^{-\frac{(x-m)^2}{2s^2}}}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}} \right) \mathrm{d}x \\
&= \int p(x) \left(\ln \left(\frac{s^2}{\sigma^2} \right)^{\frac{1}{2}} + \left(\frac{(x-m)^2}{2s^2} - \frac{(x-\mu)^2}{2\sigma^2} \right) \right) \mathrm{d}x \\
&= \ln \frac{s}{\sigma} + \frac{1}{2s^2} \int p(x)(x-m)^2 \mathrm{d}x - \frac{1}{2\sigma^2} \int p(x)(x-\mu)^2 \mathrm{d}x
\end{aligned} \tag{3.1}$$

We calculate integral of second term in (3.1)

$$\begin{aligned}
\int p(x)(x-m)^2 \mathrm{d}x &= \int p(x)(x-\mu+\mu-m)^2 \mathrm{d}x \\
&= \int (x-\mu)^2 p(x) \mathrm{d}x + \int 2(x-\mu)(\mu-m)p(x)\mathrm{d}x + \int (\mu-m)^2 p(x) \mathrm{d}x \\
&= \sigma^2 + 2(\mu-m) \int (x-\mu)p(x)\mathrm{d}x + (\mu-m)^2 \int p(x)\mathrm{d}x \\
&= \sigma^2 + (\mu-m)^2
\end{aligned} \tag{3.2}$$

Substitude equation (3.2) to (3.1), we get

$$\begin{aligned}
\mathrm{KL}(p \parallel q) &= \ln \frac{s}{\sigma} + \frac{1}{2s^2} \left(\sigma^2 + (\mu-m)^2 \right) - \frac{1}{2\sigma^2} \int p(x)(x-\mu)^2 \mathrm{d}x \\
&= \ln \frac{s}{\sigma} + \frac{1}{2s^2} \left(\sigma^2 + (\mu-m)^2 \right) - \frac{\sigma^2}{2\sigma^2} \\
&= \ln \frac{s}{\sigma} + \frac{\sigma^2 + (\mu-m)^2}{2s^2} - \frac{1}{2}
\end{aligned}$$

4 Polynomial Curve Fitting (30%)

Please use ipython to check my solution to this problem.
I use numpy, matplotlib, and pandas packages.

```
run hw1_4.py
hw1_4_1()
hw1_4_2()
hw1_4_3()
```

5 Polynomial Regression (30%)

Please use ipython to check my solution to this problem.
I use numpy, scipy and matplotlib packages.

```
run hw1_5.py
hw1_5_1()
hw1_5_2()
```