867002105_HW7

Charles Dotson

November 1, 2022

1 Q1

We start by showing the Jensen inequality. Where $f(x) = -\log x$ since it is concave for $x > 0$.

$$KL = \int p(x) \left[-\log \frac{q(x)}{p(x)} \right] dx = \mathbb{E}_p[-\log g(x)]$$

$$\geq -\log \mathbb{E}_p[g(x)] = -\log \int p(x) \frac{q(x)}{p(x)} dx = -\log \int q(x) dx = -\log(1) = 0$$

Since this must hold true, $g(x) = \frac{p(x)}{qx} = c$ but since we are able to see that if that we end up with just taking the integral of one PDF over the same domain. We can see that it must not equal a constant but actually, $g(x) = \frac{p(x)}{q(x)} = 1$

2 Q3

$$
\begin{aligned}
||x - z_1u_1 - z_2u_2||^2 &= (x - z_1u_1 - z_2u_2)^T(x - z_1u_1 - z_2u_2) \\
\\
&= (x^T - (z_1u_1)^T - (z_2u_2)^T)(x - z_1u_1 - z_2u_2) \\
\\
&= x^Tx - x^Tz_1u_1 - x^Tz_2u_2 \\
&- (z_1u_1)^Tx + < z_1u_1, z_1u_1 > + < z_1u_1, z_2u_2 > \\
&- (z_2u_2)^Tx + < z_2u_2, z_1u_1 > + < z_2u_2, z_2u_2 > \\
\\
&= x^Tx - x^Tz_1u_1 - x^Tz_2u_2 \\
&- (z_1u_1)^Tx + (x^Tu_1)^2 < u_1, u_1 > + (x^Tu_1)(x^Tu_2) < u_1, u_2 > \\
&- (z_2u_2)^Tx + (x^Tu_2)^2 < u_2, u_2 > + (x^Tu_1)(x^Tu_2) < u_1, u_2 > \\
\\
&= x^Tx - x^Tz_1u_1 - x^Tz_2u_2 \\
&- (z_1u_1)^Tx + (x^Tu_1)^2 \\
&- (z_2u_2)^Tx + (x^Tu_2)^2
\end{aligned}
$$

We know the following to be true.

$$x = u_j(u_j^Tx) = \frac{u_ju_j^T}{u_j^Tu_j}x$$

Thus we state the following.

$$\begin{aligned}
||x - z_1u_1 - z_2u_2||^2 &= x^Tx - x^T(x^Tu_1)u_1 - x^T(x^Tu_2)u_2 \\
&- ((x^Tu_1)u_1)^Tx + (x^Tu_1)^2 \\
&- ((x^Tu_2)u_2)^Tx + (x^Tu_2)^2 \\
\\
&= ||x||^2 - \frac{u_1u_1^T}{u_1^Tu_1}x\frac{u_1u_1^T}{u_1^Tu_1}x - \frac{u_2u_2^T}{u_2^Tu_2}x\frac{u_2u_2^T}{u_2^Tu_2}x \\
&- \frac{u_1u_1^T}{u_1^Tu_1}x\frac{u_1u_1^T}{u_1^Tu_1}x + (x^Tu_1)^2 \\
&- \frac{u_2u_2^T}{u_2^Tu_2}x\frac{u_2u_2^T}{u_2^Tu_2} + (x^Tu_2)^2 \\
\\
&= ||x||^2 - < u_1(u_1^Tx), u_1(u_1^Tx) > - < u_2(u_2^Tx), u_2(u_2^Tx) > \\
&- < u_1(u_1^Tx), u_1(u_1^Tx) > + (x^Tu_1)^2 \\
&- < u_2(u_2^Tx), u_2(u_2^Tx) > + (x^Tu_2)^2 \\
\\
&= ||x||^2 - (x^Tu_1)^2 < u_1, u_1 > - (x^Tu_2)^2 < u_2, u_2 > \\
&- (x^Tu_1)^2 < u_1, u_1 > + (x^Tu_1)^2 \\
&- (x^Tu_2)^2 < u_2, u_2 > + (x^Tu_2)^2 \\
\\
&= ||x||^2 - (x^Tu_1)^2 - (x^Tu_2)^2 \\
&- (x^Tu_1)^2 + (x^Tu_1)^2 \\
&- (x^Tu_2)^2 + (x^Tu_2)^2 \\
\\
&= ||x||^2 - (x^Tu_1)^2 - (x^Tu_2)^2 \\
\\
&= ||x||^2 - [(x^Tu_1)^2 + (x^Tu_2)^2] \\
\\
&= ||x||^2 - \sum_{j=1}^{2}(x^Tu_j)^2 \\
\\
&= ||x||^2 - \sum_{j=1}^{2}(z_j)^2
\end{aligned}$$

3 Q5

3.1 Commonality:

They both deal with finding a hyperplane that will seperate the classes correctly within some error.

3.2 Differences:

In LDA, the seperating hyperplane seperates the means of the samples in each class while in SVM, finds the hyperplane with the maximum of the minimum of distance between the hyperplane to the nearest point in each class.