CSC311 HW1

Jaekang Lee

February 2020

1 Nearest Neighbours and the Curse of Dimensionality

a) $E[Z] = E[X^2] - 2E[XY] + E[Y^2]$
$= E[X^2] - 2E[X]E[Y] + E[Y^2]$
$= (b^3 - a)/3(b - a) - 2((a + b)/2)^2 + (b^3 - a^3)/3(b - a)$
$= 2(a - b^3)/3(a - b) - (a + b)^2/2$
Substitute a=1, b=0, $2/3 - 1/2 = 1/6$!
$Var(Z) = E[Z^2] - E[Z]^2 = E[(x - y)^4] - (1/6)^2$
$= E[X^4] + E[Y^4] + 6E[X^2]E[Y^2] - 4E[X^3]E[Y] - 4E[X]E[Y^3] - 1/36$
$= 2/5 + 2/3 - 1 - 1/36 = 7/180$

b) $E[||X - Y||_2^2] = E[\sum_{i=1} d(x_i - y_i)^2] = dE_{x,y\sim[0,1]}[x - y)^2]$ This is because we have d many independent 'distances' of each element of random variable. Now we have $E[R] = E[Z_1] + E[Z_2] + ... + E[Z_d]$ which is just d many $E_{x,y\sim[0,1]}[(x - y)^2]$ as well therefore $E[||X - Y||_2^2] = E[R]$.
Which would be $E[((x - 1/2) - (y - 1/2))^2] = E[(x - 1/2)^2] + 2E[(x - 1/2)(y - 1/2)] + E[(y - 1/2)^2] = 2E[(x - 1/2)^2] + E[x - 1/2]E[y - 1/2] = 2E[(x - 1/2)^2] = 2\int_0^1 (x - 1/2)^2 dx = 2\int_0^1 x^2 - x + 14dx = 2(1/3 - 1/2 + 1/4) = 1/6$ then we times d $\implies d/6$
Let $Var(||X - Y||_2^2) = Q$ then $var(Q) = E[Q^2] - (E[Q])^2 = E[Q^2] - E[R]^2$
$\implies E[Q^2] = d^2 E_{x,y\sim[0,1]}[(x - y)^4]$
$var(R) = var(Z_1 + Z_2 + ... + Z_d) = E[R^2] - E[R]^2$
WTS $E[R^2] = d^2 E_{x,y\sim[0,1]}[(x - y)^4]$
We have $E[R^2]$ result in d^2 many terms and each term would be a product of a Z_i and Z_j where i,j are natural number between 1 to d. Which results in $((x - y)^2)^2 = (x - y)^4$ hence $var(R) = var(Q)$ Which would be $d^2 E(Z^2) - d^2/36$

c) $SD(||X - Y||_2^2) = \sqrt{var(||X - Y||_2^2} = \sqrt{d^2 E(Z^2) - 1/36}$. We know maximum possible squared Euclidean distance between two points within the d-dimensional unit cube $= \sqrt{d}$. Compared we can see that when mean and standard deviation are added together they come pretty close to the maximum possible SED between two points which means most points in this unit cube are distributed nearly farthest apart and approximately same distance because of small SD.

2 Information Theory

a) We know $0 \leq p(x) \leq 1$ by definition of pmf. This means $1/p(x) \geq 1$. To see this, we take $p(x) = h/g$ where g, h are any real number and $g \geq h$ since $p(x) \leq 1$, $1/p(x) = g/h$ which implies $1/p(x)$ is always greater than or equal to 1. We know $log(anything \geq 1)$ results in a real number greater than or equal to 0 hence $p(x)log(1/p(x))$ for any random variable X will return a non-negative number, hence summation over all these terms is positive.

b) We have from chain rule that $H(X,Y) = H(X) + H(Y|X)$. But since X and Y are independent, X has no impact on random variable Y so entity of Y given X is just entity of Y. Hence $H(X,Y) = H(X) + H(Y)$

c) Let $(x,y) \sim p$, we have that $p(x,y) = p_x(x) \times p_{Y|X}(x|y)$
$\implies log_2 p(x,y) = log_2 p_X(x) + log_2 p_{Y|X}(x|y)$ (log both sides)
$\implies E[log_2 p(x,y)] = E[log_2 p_X(x)] + E[log_2 p_{Y|X}(Y|X)]$ (expect value both side)
$\implies \sum_x \sum_y p(x,y) log_2(1/p(x,y)) = \sum_x p(x) log_2(1/p(x)) + \sum_x \sum_y p(x,y) log_2(1/p(y|x))$
$\implies H(X,Y) = H(X) + H(Y|X)$

d) Using Jensen's inequality, note that log function is always concave since $p(x)/q(x)$ is guarantee to be positive number because both are pmf. Recall logarithm rule where $log_2 y^z = z \times log_2 y$. Using these two rules we have the following.
$-kl(p||q) = \sum_x p(x) log_2(q(x)/p(x)) \leq log_2(\sum_x p(x) q(x)/p(x)) = log_2 1 = 0$
Explanation: p(x) cancels out then $\sum_x q(x) = 1$ by pmf definition. Hence since $-kl(p||q)$ is always negative, it follows that $+kl(p||q)$ is always non-negative.

e) $I(Y;X) = H(Y) - H(Y|X) = KL(p(x,y)||p(x)p(y))$
$\implies \sum_{(x,y)} p(x,y) log_2(p(x)p(y|x)/p(x)p(y))$
$\implies -\sum_{(x,y)} p(x,y) log_2 p(y) + \sum_{(x,y)} log_2 p(y|x)$
$\implies -\sum_y p(y) log_2 p(y) + \sum_y p(x) log_2 H(Y|X=x)$
$\implies H(Y) - H(Y|X)$

3 Question 3