Problem - 1 :

*)
$$pdf = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\right\}$$

$$\Sigma = \sigma^2 I$$

$$\det|\Sigma| = (\sigma^2)^d$$

$$\Sigma^{-1} = \frac{1}{\sigma^2} I$$

$$\Rightarrow pdf = \frac{1}{(2\pi)^{d/2}(\sigma^2)^{d/2}} \times \exp\left\{-\frac{1}{2} x \frac{1}{\sigma^2} I x^T\right\}$$

$$\Rightarrow pdf = \frac{1}{(2\pi\sigma^2)^{d/2}} \times \exp\left\{\frac{-xx^T}{2\sigma^2}\right\}$$

$$xx^T = \sum_{i=1}^{d} x_i^2 = r^2$$

$$\Rightarrow pdf = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left\{\frac{-r^2}{2\sigma^2}\right\}$$

Volume of hyper sphere $= C r^n$

Surface Area $(S) = \frac{dV}{dr} = nCr^{n-1}$

$$S_d = nC$$

$$\Rightarrow S = S_d r^{n-1} = S_d r^{d-1}$$

$$\int pdf \; dx_1 \, dx_2 \cdots dx_d$$

$$= \int \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-r^2/2\sigma^2} \, dx_1 \cdots dx_d$$

$$= \frac{1}{(2\pi\sigma^2)^{d/2}} \times e^{-r^2/2\sigma^2} \times \int dx_1 \cdots dx_d$$

$$= \frac{1}{(2\pi\sigma^2)^{d/2}} \times e^{-r^2/2\sigma^2} \times \overbrace{S \times \epsilon}^{\text{Volume}}$$

$$= \frac{1}{(2\pi\sigma^2)^{d/2}} \times e^{-r^2/2\sigma^2} \times S_d \, r^{d-1} \times \epsilon$$

$$= \underline{\underline{\frac{S_d \, r^{d-1}}{(2\pi\sigma^2)^{d/2}} e^{-r^2/2\sigma^2} \epsilon .}}$$

Problem-1

2.)

Stationary point $\Rightarrow \frac{df}{dx} = 0$

$$p(r/d) = \frac{S_d\, r^{d-1}}{(2\pi\sigma^2)^{d/2}} e^{-r^2/2\sigma^2}$$

$$\frac{S_d}{(2\pi\sigma^2)^{d/2}} = K \quad \text{(Constant w.r.t to } r\text{)}$$

$$\frac{d\left(K r^{d-1} e^{-r^2/2\sigma^2}\right)}{dr} = 0$$

$$K\left[(d-1) r^{d-2} e^{-r^2/2\sigma^2} + r^{d-1} e^{-r^2/2\sigma^2} \cdot \frac{-\cancel{2}r}{\cancel{2}\sigma^2}\right] = 0$$

$$\Rightarrow K r^{d-2}\left[(d-1) e^{-r^2/2\sigma^2} - \frac{r^2 e^{-r^2/2\sigma^2}}{\sigma^2}\right] = 0$$

$$\Rightarrow K r^{d-2} e^{-r^2/2\sigma^2}\left[(d-1) - \frac{r^2}{\sigma^2}\right] = 0$$

$$\Rightarrow r \approx \sqrt{d}\,\sigma \quad (\sqrt{d-1}\,\sigma)$$

(as 'd' is large $\sqrt{d-1} \approx \sqrt{d}$)

Problem - 1

4.)

$$\text{pdf at origin} = \frac{e^{-\frac{0^2}{2\sigma^2}}}{(2\pi\sigma^2)^{d/2}}$$

$$= \frac{1}{(2\pi\sigma^2)^{d/2}}$$

$$p(r/d)\epsilon = \frac{S_d\, r^{d-1}}{(2\pi\sigma^2)^{d/2}} e^{-\frac{r^2}{2\sigma^2}} \epsilon$$

we can ignore 'ϵ' as it is constant and (ϵ is the thickness of the shell)

$$p(r/d) = \frac{S_d\, r^{d-1}}{(2\pi\sigma^2)^{d/2}} e^{-\frac{r^2}{2\sigma^2}}$$

$$\frac{S_d}{(2\pi\sigma^2)^{d/2}} = K$$

$$p(r/d) = K\, r^{d-1} e^{-\frac{r^2}{2\sigma^2}}$$

$$\frac{d(p(r/d))}{d(r)} = 0 \quad \text{at} \quad r = \sqrt{d-1}\,\sigma$$

$$\frac{d}{dr}\left[K r^{d-2} e^{-r^2/2\sigma^2}\left(d-1-\frac{r^2}{\sigma^2}\right)\right]$$

$$= K\left[\left((d-2) r^{d-3} e^{-r^2/2\sigma^2} + r^{d-2} e^{-r^2/2\sigma^2}\left(\frac{-2r}{2\sigma^2}\right)\right)\right.$$
$$\left.\left(d-1-\frac{r^2}{\sigma^2}\right) + r^{d-2} e^{-r^2/2\sigma^2}\left(-\frac{2r}{\sigma^2}\right)\right]$$

$$= K r^{d-3} e^{-r^2/2\sigma^2}\left[\left((d-2) + \frac{r(-r)}{\sigma^2}\right)\left(d-1-\frac{r^2}{\sigma^2}\right)\right.$$
$$\left. + \left(-\frac{2r^2}{\sigma^2}\right)\right]$$

$$\frac{r^2}{\sigma^2} = d-1$$

$$= K r^{d-3} e^{-\frac{r^2}{2\sigma^2}}\left[\left((d-2)-(d-1)\right)\left(d-1 - (d-1)\right)\right.$$
$$\left. - 2(d-1)\right].$$

$$= K r^{d-3} e^{-\frac{(d-1)}{2}}\left(-2(d-1)\right) < 0$$

for all $d > 1$

$$\Rightarrow \frac{d^2 p(r/d)}{dr^2} < 0.$$

max probability mass in a thin shell $= \dfrac{S_d\, r^{d-1}}{(2\pi\sigma^2)^{d/2}}\, e^{-\frac{r^2}{2\sigma^2}} \cdot \epsilon$

$$= \frac{S_d\, r^{d-1}}{(2\pi\sigma^2)^{d/2}}\, e^{-\frac{r^2}{2\sigma^2}}\, \epsilon$$

$$= \frac{S_d\, (\sqrt{d-1})^{d-1}\, \sigma^{d-1}}{(2\pi)^{d/2} \cdot \sigma^d} \cdot e^{-\frac{d-1}{2}} \cdot \epsilon$$

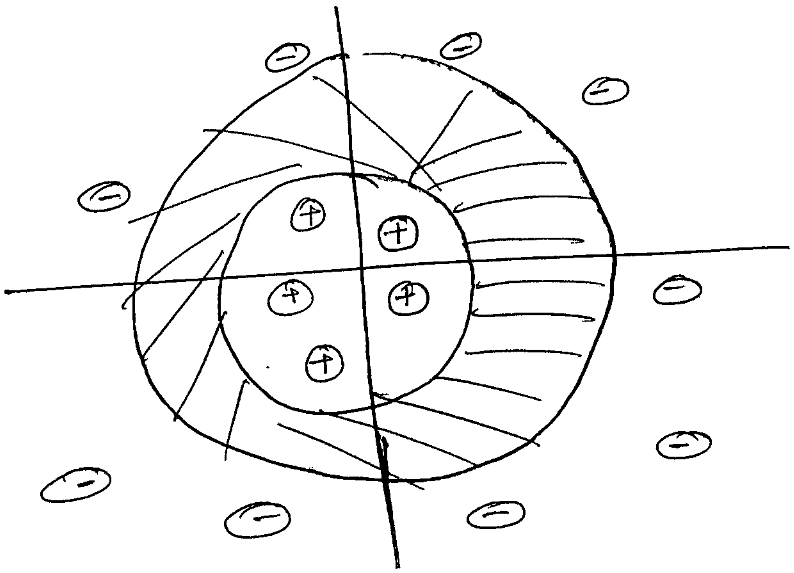
$$\frac{1}{(2\pi\sigma^2)^{d/2}} = \text{pdf at origin } (P_0)$$

$$= P_0\, S_d\, (\sqrt{d-1})^{d-1} \cdot \sigma^{d-1} \cdot e^{-\left(\frac{d-1}{2}\right)} \cdot \epsilon$$

$$= P_0 \left(S_d\, \epsilon \left(\sqrt{\frac{d-1}{e}} \cdot \sigma \right)^{d-1} \right)$$

Problem 2:

ε, δ



Probability to miss the strip is $(1-\varepsilon)$

Probability of 'N' points to miss is $(1-\varepsilon)^N$

$$(1-\varepsilon)^N \le \delta$$

$$(1-x) \le \exp(-x)$$

$$\Rightarrow [\exp(-\varepsilon)]^N \le \delta$$

$$\Rightarrow \exp(-\varepsilon N) \le \delta$$

$$\Rightarrow -\varepsilon N \le \log\delta$$

(log is increasing function)

$$\Rightarrow \varepsilon N \ge \log(1/\delta)$$

$$\Rightarrow N \ge \frac{1}{\varepsilon}\log(1/\delta).$$

Problem 3

$$E\left[L_q(t, y(x))\right] = \int_{t,x} |t - y(x)|^q \, p(t,x) \, dt \, dx$$

1.) for $q = 2$:

$$\frac{\partial E}{\partial y} = 0 \quad \& \quad \frac{\partial^2 E}{\partial y^2} > 0 \rightarrow \quad -2(-1) \int_{t,x} p(t,x) \, dt \, dx = \underline{\underline{2 > 0}}$$

$$-2 \int_{t,x} (t - y(x)) \, p(t,x) \, dt \, dx = 0$$

$$\Rightarrow \int_t \int_x (t - y(x)) \, p(t,x) \, dt \, dx = 0$$

$$\Rightarrow \int_t \int_x t \, p(t,x) \, dt \, dx = \int_t \int_x y(x) \, p(t,x) \, dt \, dx$$

$$\int_x \int_t t \, p(t|x) \, p(x) \, dt \, dx = \int_x \int_t y(x) \, p(t,x) \, dt \, dx$$

$$\int_x p(x) \left(\int_t t \, p(t|x) \, dt \right) dx = \int_x y(x) \left(\int_t p(t,x) \, dt \right) dx$$

$$\int_x \left(\int_t t \, p(t|x) \, dt \right) p(x) \, dx = \int_x y(x) \, p(x) \, dx$$

$$\Rightarrow \quad y(x) = \int_t t\, p(t/x)\, dt$$

is a solution

$$\Rightarrow \quad y(x) = E[t/x]$$

2.) for $q=1$:

$$E[L_q(t, y(x))] = \int_{t,x} \underbrace{|t - y(x)|^1}_{\text{piecewise linear curve}} p(t,x)\, dt\, dx$$

(not differentiable at $t=y$)

$$\frac{\partial E}{\partial y} = -\int_{t,x} \frac{|t-y(x)|}{t-y(x)} p(t,x)\, dt\, dx = 0$$

for $t > y(x)$ for $t < y(x)$

$$\Rightarrow \left(\int_{t,x} p(t,x)\, dt\, dx = \left(\int_{t,x} p(t,x)\, dt\, dx\right.\right.$$

(only $t \neq y(x)$ cases contribute to Loss.)

$$\Rightarrow \quad P(t > y(x)) = P(t < y(x))$$

$\left(\text{as } \dfrac{\partial^2 E}{\partial y^2} > 0\right) \Rightarrow$ minimum value occurs at $y = \text{median}(t/x)$

3.) for $q \to 0$

$$\underset{q \to 0}{Lt} \int_{t,x} |t - y(x)|^q p(t, x)\, dt\, dx$$

$$\underset{q \to 0}{Lt} (a)^q = 1$$ for all 'a' except for $\underline{\underline{a=0}}$

$$\Rightarrow$$

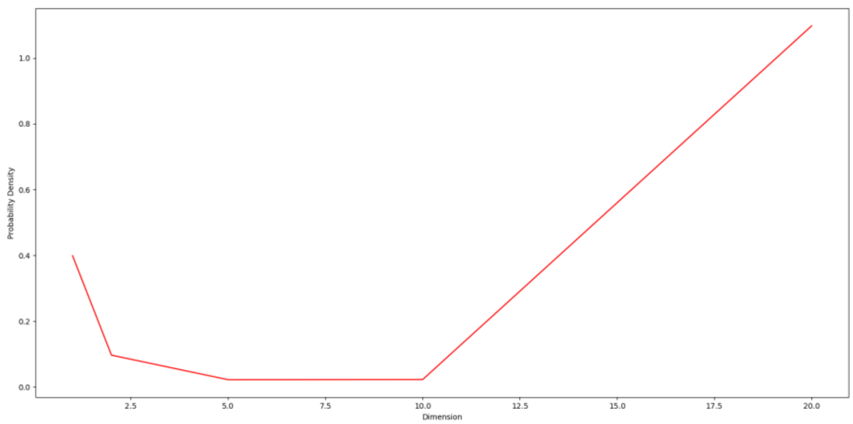
$$= \underset{q \to 0}{Lt} \int_{t,x} |t - y(x)|^q p(t, x)\, dt\, dx$$

$$= \underset{q \to 0}{Lt} \int_x \left(\int_t |t - y(x)|^q p(t/x)\, dt \right) p(x)\, dx$$

the more times $t - y(x) = 0$, the less the loss value will be,

So, for a given 'x', 'y' should be the mode ~~median~~ of t.

$$\Rightarrow \quad y^*(x) = \text{mode}(t/x)$$



Problem -1, Question -3 plot

Question 5.1.2:

Train ratio	1	3	5	7	9	11	15	19	K Selection
0.1	82.50	85.0	85.0	85.0	**90.0**	85.0	**90.0**	80.0	15
0.2	87.65	87.65	86.42	82.72	83.95	87.65	**90.12**	88.88	15
0.3	88.52	88.52	**90.16**	86.88	86.88	87.70	88.52	**90.16**	19
0.4	90.74	90.74	**91.36**	90.12	88.27	90.12	87.04	88.89	5
0.5	**93.60**	91.62	92.11	91.13	91.13	90.64	90.15	91.13	5
0.6	**91.80**	90.57	**91.80**	90.98	90.57	89.34	88.11	89.75	5
0.7	**92.25**	**92.25**	91.90	91.19	89.79	89.79	88.73	87.68	3
0.8	92.30	92.00	**92.62**	91.08	89.54	89.54	90.46	88.62	5
0.9	**93.72**	92.90	92.34	91.26	90.44	89.62	90.44	90.16	5
1.0	**94.34**	93.34	92.36	92.61	92.12	91.63	90.89	91.87	5

Rule: I pick the k for which the accuracy is high, at the same time I avoided picking k values which are small, as they might lead to overfitting or vulnerable noise or some irrelevant features. In that cases I pick the next best k-values.

For ratio = 0.1, I got maximum validation accuracy for k = 9,15. But I choose 15 because it gives smoother decision boundary and robust to noise.

For ratio = 0.2, I got maximum validation accuracy at k = 15, I choose 15 as the k.

For ratio = 0 .3, I got maximum validation accuracy at k = 5,19. But I choose 19 because it gives smoother decision boundary and robust to noise.

For ratio = 0 .4, I got maximum validation accuracy at k = 5, I choose 5 as the k.

For ratio = 0 .5, I got maximum validation accuracy at k = 1. But I choose the next best because k=5 is vulnerable to noise and irrelevant features. (irrelevant features this can be avoided by using weighted

Euclidean distance.)
Also the difference is just 1.5%.

For ratio = 0 .6, I got maximum validation accuracy at k = 1,5. But I choose 5 because it gives smoother decision boundary and robust to noise.

For ratio = 0 .7, I got maximum validation accuracy at k = 1,3. But I choose 5 because the difference between the accuracies is very less (0.35%) and k=5 gives smoother decision boundary and robust to noise.

For ratio = 0 .8, I got maximum validation accuracy at k = 5, I choose 5 as the k.

For ratio = 0 .9, I got maximum validation accuracy at k = 1, I choose 5 as difference between the accuracies is very less (1.4 %)and k=5 gives smoother decision boundary and robust to noise.

For ratio = 1.0, I got maximum validation accuracy at k = 1, I choose 5 as the k. I choose 5 as difference between the accuracies is very less (1.9 %)and k=5 gives smoother decision boundary and robust to noise.

Question 5.1.3

Train ratio	1	3	5	7	9	11	15	19	K Selection
0.1	75.38	75.46	77.01	76.86	77.08	76.34	76.49	**77.30**	15
0.2	77.53	78.56	79.89	80.63	81.89	**82.63**	**82.63**	81.89	15
0.3	79.08	80.26	80.56	82.33	82.55	83.44	82.63	**84.10**	19
0.4	80.48	81.37	81.30	82.63	84.10	83.59	83.37	**84.25**	19
0.5	81.22	81.15	81.44	82.55	83.88	83.66	83.88	**84.47**	19
0.6	81.30	81.74	81.37	82.85	84.03	83.66	84.10	**84.92**	19
0.7	82.55	83.14	82.92	84.47	84.18	84.18	84.10	**84.55**	19
0.8	82.77	83.66	83.29	84.10	83.88	**85.07**	84.40	84.40	15
0.9	83.66	84.77	84.47	83.73	84.10	**85.36**	84.99	84.77	5
1.0	83.96	84.62	85.36	84.77	84.77	85.07	**86.10**	85.66	19

Rule: We pick the k for which the test accuracy value is high, if the k is same for test and validation then we pick that k, else we will select that k for which the average of test and validation accuracy is high. We avoid picking those k's for which the difference between test and validation accuracy is high.

For ratio = 0.1, I got maximum accuracy at k = 19, But I choose k = 15 because as it has high validation accuracy when compared to k=19 model, also k=15 test accuracy is less than k = 19 by 0.80%.

For ratio = 0.2, I got maximum accuracy at k = 11,15. So, I choose 15 as the k because for k=15 validation accuracy is better when compared with validation accuracy of k = 11.

For ratio = 0.3, I got maximum accuracy at k = 19, I choose 19 as the k.

For ratio = 0.4, I got maximum accuracy at k = 19, I choose 19 as the k. (Average accuracies are close to each other, so I preferred larger k-value that is large because it will robust to noise)

For ratio = 0.5, I got maximum accuracy at k = 19, I choose 19 as the k. (Average accuracies are close to each other, so I preferred larger k-value that is large because it will robust to noise)

For ratio = 0.6, I got maximum accuracy at k = 19, I choose 19 as the k. (Average accuracies are close to each other, so I preferred larger k-value that is large because it will robust to noise)

For ratio = 0.7, I got maximum accuracy at k = 19, I choose 19 as the k. (Average accuracies are close to each other, so I preferred larger k-value that is large because it will robust to noise)

For ratio = 0.8, I got maximum accuracy at k = 11, I choose 15 as the k. (Average of k = 7 and k= 15 are close, so I pick up the larger k)

For ratio = 0.9, I got maximum accuracy at k = 11, I choose 5 as the k. (k=5 has better average accuracy)

For ratio = 1.0, I got maximum accuracy at k = 15, I choose 19 as the k.

Question 5.1.4:

As the number of data samples increases, we can see that the best k value is high, this is because as we increase the dataset size our train data distribution tends to true distribution. The effect of outliers decreases with the increase of value of k. In general, we can say that, irrespective of the k-size the value of accuracy increased (on average).

Instead of validation we did, we can do cross validation (this more unbiased than validation, as validation may depend on how we split the data). We can also eliminate outliers from the train data. (by eliminating the ones which doesn’t fit the model trained on it). We can do PCA or any other feature reduction technique to select the most relevant features by which we can improve the accuracy of the whole system. We should also try to avoid low k values. (low k-values might lead to overfitting)