Problem 1:

In Naive Bayes if σ is constrained to be diagonal then $P(X^j|Y)$ can be written as a product of $P(X^j|Y)$

$$P(X|Y) = \frac{1}{\sqrt{(2\pi)^D \det(\sigma_Y)}} \exp\left(-\frac{1}{2}(X^j - \mu^j_{jY})^T \sigma_Y^{-1} (X_K - \mu_{KY})\right)$$

$$= \prod_{j=1}^{D} \frac{1}{\sqrt{(2\pi)^D \sigma_{Yjj}}} \exp\left(-\frac{1}{2\sigma_{Yjj}} \|X^j - \mu_{jY}\|_2^2\right) = \prod_{j=1}^{D} P(X^j|Y)$$

So diagonal covariance matrix satisfies naive bayes assum

Case 1: The covariance matrix is shared.

$P(X|Y) = N(X|\mu_y, \sigma)$

$P(X|Y=1) = P(X|Y=0)$

$\log \pi_1 - \frac{1}{2}(X-\mu_1)^T \sigma^{-1} (X-\mu_1) = \log \pi_0 - \frac{1}{2}(X-\mu_0)^T \sigma^{-1} (X-\mu_0)$

$c + X^T\sigma^{-1}X - 2\mu_1^T\sigma^{-1}X + \mu_1^T\sigma^{-1}\mu_1 = X^T\sigma^{-1}X - 2\mu_0^T\sigma^{-1}X + \mu_0^T\sigma^{-1}\mu_0$

$[2(\mu_0-\mu_1)^T\sigma^{-1}] - X(\mu_0-\mu_1)^T\sigma^{-1}(\mu_0-\mu_1) = c$

$\Rightarrow b^i x^i + c = 0$

where, $b^i = 2(\mu_0-\mu_1)^T\sigma^{-1}$

$c = -(\mu_0-\mu_1)^T\sigma^{-1}(\mu_0-\mu_1)$

This is a linear function

If the covariance is not shared.

$$x^T(\sigma_1^{-1}-\sigma_0^{-1})x-2(\mu_1^T\sigma_1^{-1}-\mu_0^T\sigma_0^{-1})x+(\mu_0^T\sigma_0\mu_0-\mu_1^T\sigma_1\mu_1)=C$$

$$\Rightarrow x^T a_i x + b_i x + c_i = 0$$

where. $a_i = \sigma_1^{-1} - \sigma_0^{-1}$

$b_i = -2(\mu_1^T\sigma_1^{-1} - \mu_0^T\sigma_0^{-1})$

$c_i = (\mu_0^T\sigma_0\mu_0 - \mu_1^T\sigma_1\mu_1)$

This is a quadratic function.

Problem 2

(a) $S_{m-1}(r)$

$= S_{2-1}(r)$ $[m=2]$

$= S_1(r)$

$= 2\pi r$

$V_m(r)$

$= V_2(r)$

$= \pi r^2$

$S_{m-1}(r)$

$= S_2(r)$ $[m=3]$

$= 4\pi r^2$

$V_m(r)$

$= V_3(r)$

$= \frac{4}{3}\pi r^3$

$V_0 = 1$

$V_1 = 2$

$S_0 = 2$

$S_1 = 2\pi$

$V_2 = \pi$

$S_2 = 4\pi$

$V_3 = \frac{4}{3}\pi$

(b) $V = \frac{4}{3}\pi r^3$

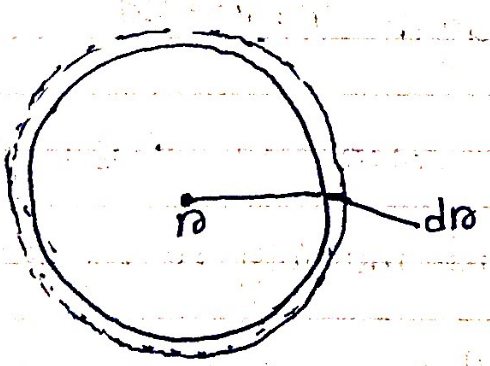
$\frac{d}{dx}[x^n] = nx^{n-1}$

$\frac{dv}{dr} = V'(r) = \frac{4}{3}\pi 3r^2$

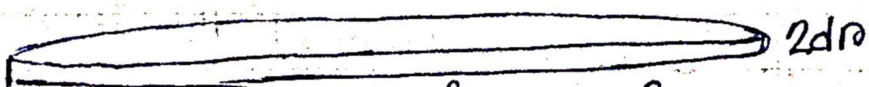
$\frac{dv}{dr} = V'(r) = 4\pi r^2$

$S = 4\pi r^2$

$\frac{dv}{dr}$ describes the change in volume with respect to the radius



A small increase in radius increase the volume of the circle
If we flatten this shell or layer created by the increase
radius it will produce a right circular cylinder



$A = \frac{4\pi r^2}{2} = 2\pi r^2$

Suppose a ball filled with air represent the shell or layer we want to find the volume of. If we begin to squish that and make that completely flattened it will look like a right circular cylinder. (in the previous page) which represents:

Top circle thickness $= dr$

Bottom " " $= dr$

Total thickness or height $= dr + dr = 2dr$

In order to find the volume we first calculate area of the base and multiply it by the height.

Area of the base of the right circular cylinder

$A_{base} = \frac{4\pi r^2}{2} = 2\pi r^2$ [half of the surface of the sphere]

So, $V = A_{base} \cdot h = 2\pi r^2 \cdot 2dr$

This is the changed volume of the sphere

So, $V' = 4\pi r^2 dr$

$$\frac{dV}{dr} = 4\pi r^2$$

So the derivative of the volume of a sphere is the surface area. Increasing the radius results in an increase in volume proportional to the surface area

(c) $S_{m-1}(r) = \bar{S}_{m-1} r^{m-1}$

$$V_m = \frac{\bar{S}_{m-1}}{m}, \text{ therefore } V_{m-2} = \frac{\bar{S}_{m-3}}{m-2} \quad \text{---- (1)}$$

$$\bar{S}_{m-1} = 2\pi V_{m-2}$$

$$= \frac{2\pi \bar{S}_{m-3}}{m-2} \text{ [From eqn (1)]}$$

$$\boxed{\text{So, } S_{m-1}(r) = \frac{2\pi \bar{S}_{m-3}}{m-2} r^{m-1}}$$

2(d)

$$P(x) = \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{\|x^2\|}{2\sigma^2}\right)$$

$$S_{m-1}(r) = \frac{2\pi S_{m-3}}{m-2} r^{m-1}$$

$$x \in S_{m-1}(r)$$

$$P_m(r) = \int_{x \in S_{m-1}(r)} P(x)\,dx$$

Things we need to do:

1. we have to chop up the region $S_{m-1}(r)$ into tiny pieces
2. Multiply the area of each piece dx, by the value of $P(x)$ at one of the points inside that piece
3. Add up the resulting values

So, $$P_m(r) = \int_{S_{m-1}(r)} \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{\|x^2\|}{2\sigma^2}\right) dx$$

But since P depends only on x as well as r, it is constant on the spherical surface

So the equation $\rho_m(r)$ for the integrated density of sampled points from the gaussian distribution lying on the surface of $S_{m-1}(r)$:

$$\boxed{\rho_m(r) = r^{m-1} e^{-\frac{r^2}{2\sigma^2}}}$$

$\boxed{2(e)}$ $\rho_m(r)$ will be maximum when

$$\frac{d\rho_m(r)}{dr} = 0$$

$$\Rightarrow e^{-\frac{r^2}{2\sigma^2}} (m-1) r^{m-1-1} - r^{m-1} e^{-\frac{r^2}{2\sigma^2}} \cdot \frac{2r}{2\sigma^2} = 0$$

$$\Rightarrow e^{-\frac{r^2}{2\sigma^2}} (m-1) r^{(m-2)} - \frac{r^m}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} = 0$$

$$\Rightarrow e^{-\frac{r^2}{2\sigma^2}} \left((m-1) r^{m-2} - \frac{r^m}{\sigma^2} \right) = 0$$

$$\Rightarrow (m-1) r^{m-2} = \frac{r^m}{\sigma^2}$$

$$\Rightarrow r^{m-2-m} = \frac{1}{(m-1)\sigma^2}$$

$$\Rightarrow r^{-2} = \{(m-1)\sigma^2\}^{-1}$$

$$\Rightarrow r^2 = (m-1)\sigma^2$$

As m is very large

$m-1 \approx m$

so, $\boxed{\hat{r} = \sqrt{m}\,\sigma}$

$\boxed{2(f)}$ $P(r) = r^{m-1} e^{-\frac{r^2}{2\sigma^2}}$

$\Rightarrow P(\hat{r}+\varepsilon) = (\hat{r}+\varepsilon)^{m-1} e^{\frac{-(\hat{r}+\varepsilon)^2}{2\sigma^2}}$

$\ln P(\hat{r}+\varepsilon) = (m-1)\ln(\hat{r}+\varepsilon) - \frac{(\hat{r}+\varepsilon)^2}{2\sigma^2} = S(\hat{r}+\varepsilon)$

Differentiating

$S'(\hat{r}+\varepsilon) = \frac{m-1}{\hat{r}+\varepsilon} - \frac{\hat{r}+\varepsilon}{\sigma^2}$

$S''(\hat{r}+\varepsilon) = -\frac{(m-1)}{(\hat{r}+\varepsilon)^2} - \frac{1}{\sigma^2} \leq -1$

$S'(\hat{r}+\varepsilon) = 0$ at $(\hat{r}+\varepsilon)^2 = (m-1)\sigma^2$

$\Rightarrow \hat{r}+\varepsilon = \sqrt{m}\,\sigma = \hat{r}$

$S''(r+\varepsilon) < 0$ for all r.

The taylor series expansion for $S(\hat{r}+\varepsilon)$ is.

$$S(\hat{n}+\varepsilon) = S(\hat{n}) + S'(\hat{n})(\hat{n}+\varepsilon-\hat{n}) + \frac{1}{2}S''(\hat{n})(\hat{n}+\varepsilon-\hat{n})^2 + \cdots$$

Thus.

$$S(\hat{n}+\varepsilon) = S(\hat{n}) + S'(\hat{n})(\hat{n}+\varepsilon-\hat{n}) + \frac{1}{2}S''(\xi)(\hat{n}+\varepsilon-\hat{n})^2$$

For some point ξ between $\hat{n}$ and $\hat{n}+\varepsilon$ since $S'(\hat{n}) = 0$, the second term vanishes

$$S(\hat{n}+\varepsilon) = S(\hat{n}) + \frac{1}{2}S''(\xi)(\hat{n}+\varepsilon-\hat{n})^2$$

Since the second derivative is always less than -1,

$$S(\hat{n}+\varepsilon) \leq S(\hat{n}) - \frac{1}{2}(\hat{n}+\varepsilon-\hat{n})^2$$

now $P(\hat{n}+\varepsilon) = e^{S(\hat{n}+\varepsilon)}$ Therefore

$$P(\hat{n}+\varepsilon) \leq e^{S(\hat{n}) - \frac{1}{6^2}(\hat{n}+\varepsilon-\hat{n})^2}$$

$$\therefore \boxed{P(\hat{n}+\varepsilon) \leq P(\hat{n})e^{-\varepsilon^2/6^2}}$$

2g

For a low dimension gaussian distribution most points are close to origin. For a high dimension gaussian distribution the mass is concentrated in a thin annulus of width $O(1)$ at radius $\sqrt{m}$, located outside the sphere.

2h

Probability density at origin:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Probability density at one point on sphere $S_{m-1}(\hat{r})$

$$P(x) = \frac{1}{(2\pi)^{d/2}\sigma^d} \exp\left(-\frac{|x|^2}{2\sigma^2}\right)$$

one way to illustrate the vastness of high-dimensiona space is comparing the volume:

The volume of a hypersphere:

$$\frac{2r^d \pi^{d/2}}{d\Gamma(d/2)}$$

when the volume of a hypercube:

$2r$.

Comparing the proportions.

$$\frac{V_{hypersphere}}{V_{hypercube}} = \frac{\pi^{d/2}}{d\,2^{d-1}\,\Gamma(d/2)} \to 0 \text{ as } d \to \infty$$