1. (a) Not a kernel; kernels are scalar functions.

(b) $K(\underline{x}, \underline{x}') = \left(f(\underline{x}) + g(\underline{x})\right)\left(f(\underline{x}') + g(\underline{x}')\right)$

$$= \underline{\phi}(\underline{x})^T \underline{\phi}(\underline{x}')$$

where $\underline{\phi}(\underline{x}) \triangleq \left(f(\underline{x}) + g(\underline{x})\right)$

$\Rightarrow K$ is a kernel.

(c) $$e^{\frac{-\|\underline{x} - \underline{x}'\|^2}{\sigma^2}} = e^{\frac{-\|\underline{x}\|^2 - \|\underline{x}'\|^2 + 2\underline{x}^T\underline{x}'}{\sigma^2}}$$

$$= e^{-\frac{\|\underline{x}\|^2}{\sigma^2}} e^{-\frac{\|\underline{x}'\|^2}{\sigma^2}} e^{\frac{2\underline{x}^T\underline{x}'}{\sigma^2}}$$

Consider the last term: supposing we define

$$K_1(\underline{x}, \underline{x}') = \frac{2\underline{x}^T\underline{x}'}{\sigma^2}$$

This is clearly a valid kernel (positive constant times dot product)

So we have

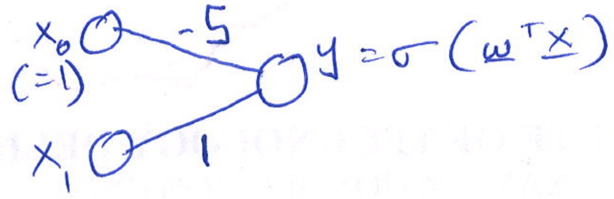
$$e^{K_1(\underline{x}, \underline{x}')} = \lim_{i \to \infty} \left(1 + K_1 + \frac{K_1^2}{2} + \dots + \frac{K_1^i}{i!}\right)$$

we can show that any positive-coefficient polynomial fn. of a kernel is also a kernel (see soln. 2.(h) of Problem Set 3).

Also, if we define $\underline{\psi}(\underline{x}) = \left(e^{-\frac{\|\underline{x}\|^2}{\sigma^2}}\right)$, we see that $e^{\frac{-\|\underline{x}\|}{\sigma^2}} e^{\frac{-\|\underline{x}'\|^2}{\sigma^2}}$ is a kernel.

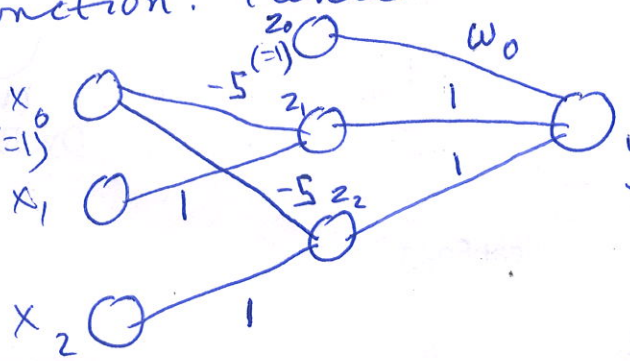
Hence K, which is the product of these two, is (by soln. 2.(e) of P.S. 3) a valid kernel.

2. (a)



Dec. boundary is $1 \cdot x_1 - 5 \cdot x_0 = 0$, or $\underline{x_1 = 5}$

(b) we want two hidden units like the above, for $x_1 = 5$ and $x_2 = 5$. Then we should OR their outputs, to get the desired function. Hence:

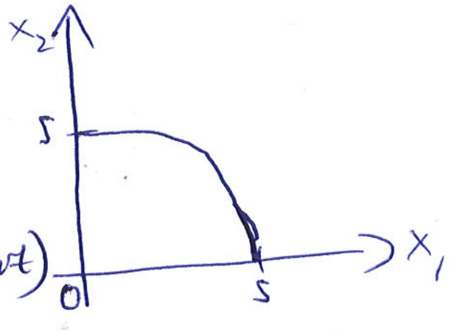


$y = \sigma(w_0 z_0 + z_1 + z_2)$

Decision boundary:

$\underline{w_0 + z_1 + z_2 = 0}$

If we assume z_1 and z_2 fire only when $x_1 > 5$ and $x_2 > 5$ respectively, then we could infer that any $w_0 \in [-1, 0]$ would give us the required boundary. Actually, due to the the activation is soft, sigmoid, so this is not strictly true. But the natural choice for w_0 seems to be in the middle of the range, i.e., $\underline{w_0 = -0.5}$ With this you will get the following boundary:



(the sharp corner will be smoothed out)

3. $$\frac{\partial E}{\partial y_n} = -\frac{t_n}{y_n} + \frac{1-t_n}{1-y_n}$$

Set this $= 0$:

$$\frac{-t_n + \cancel{t_n y_n} + y_n - \cancel{t_n y_n}}{y_n(1-y_n)} = 0$$

$\Rightarrow$ $\underline{y_n = t_n}$ gives the minimum

[Clearly E has no maximum; $E \to \infty$ as $y_n \to 0/1$]

$$E_{min} = -\sum_n t_n \log t_n + (1-t_n)\log(1-t_n)$$

($E_{min} = 0$ <u>only</u> if $t_n \in \{0,1\}\ \forall n$)

4. (a) Notation:

Weight $- x_1$

Sex $- x_2$ $\left(\begin{array}{l} x_2 = 1 \text{ for M} \\ x_2 = 0 \text{ for F} \end{array}\right)$

Has Diabetes $- t$ $\left(\begin{array}{l} t = 1 \text{ for Y} \\ t = 0 \text{ for N} \end{array}\right)$

Joint distr. factors as:

$$p(\underline{x}, t) = p(x_1|t) \cdot p(x_2|t) \cdot p(t)$$

$p(t)$: Bernoulli distr.; let us denote

$$p(t=1) = \boxed{\pi}$$

$p(x_1|t)$: Gaussians; let us denote

$p(x_1 | t=1) \sim \mathcal{N}(\boxed{\mu_1}, \boxed{\sigma_1^2})$

$p(x_1 | t=0) \sim \mathcal{N}(\boxed{\mu_0}, \boxed{\sigma_0^2})$

$p(x_2 | t)$: Bernoullis ; let us denote

$p(x_2 = 1 | t=1) = \boxed{p_1}$

$p(x_2 = 1 | t=0) = \boxed{p_0}$

So total of 7 parameters (boxed above).

(b) Using index 'n' over data points : [n^{th} point $= (\underline{x}_n, t_n)$]

$$\hat{\pi}_{ML} = \frac{\sum_{n=1}^{N} t_n}{N} \qquad \hat{\mu}_{1_{ML}} = \frac{1}{\sum_n t_n} \sum_n t_n x_{n1}$$

$$\hat{\mu}_{0_{ML}} = \frac{1}{\sum_n (1-t_n)} \sum_n (1-t_n) x_{n1} \qquad \hat{\sigma}^2_{1_{ML}} = \frac{1}{\sum_n t_n} \sum_n t_n (x_{n1} - \hat{\mu}_{1_{ML}})^2$$

$$\hat{\sigma}^2_{0_{ML}} = \frac{1}{\sum_n (1-t_n)} \sum_n (1-t_n)(x_{n1} - \hat{\mu}_{0_{ML}})^2$$

$$\hat{p}_{1_{ML}} = \frac{1}{\sum_n t_n} \sum_n t_n x_{n2} \qquad \hat{p}_{0_{ML}} = \frac{1}{\sum_n (1-t_n)} \sum_n (1-t_n) x_{n2}$$

Plug in values:

$$\hat{\pi}_{ML} = \frac{4}{8} = 0.5 \qquad \hat{\mu}_{1_{ML}} = \frac{82.7 + 88.3 + 79.7 + 83.1}{4}$$

$$= 83.45$$

$$\hat{\mu}_{0_{ML}} = \frac{67.4 + 72.7 + 70.3 + 78.4}{4} = 72.2$$

$$\hat{\sigma}^2_{1_{ML}} = \frac{(82.7 - 83.45)^2 + (88.3 - 83.45)^2 + (79.7 - 83.45)^2 + (\cancel{78.4}\,83.1 - 83.45)^2}{4}$$

$$= 9.57$$

$$\hat{\sigma}^2_{0_{ML}} = \frac{(67.4 - 72.2)^2 + (72.7 - 72.2)^2 + (70.3 - 72.2)^2 + (78.4 - 72.2)^2}{4}$$

$$= 16.34$$

$$\hat{p}_{1_{ML}} = \frac{2}{4} = 0.5 \qquad \hat{p}_{0_{ML}} = \frac{2}{4} = 0.5$$

(c) Weight feature (x_1) : The two class-conditional distributions have quite different means. Even if we look at $\mu \pm \sigma$, we have

$$83.45 \pm 3.09 \text{ and } 72.2 \pm 4.04$$

so completely non-overlapping. Hence this feature will be useful in distinguishing.

Sex feature (x_2) : The two class-conditional distributions are exactly the same : $\hat{p}_{1_{ML}} = \hat{p}_{0_{ML}}$. Hence not of any use in classification.