$\boxed{3}$ Given, k_1 and k_2 be valid kernel functions.

(a) $k(x, z) = k_1(x, z) + k_2(x, z)$

Proof: k_1 has its feature map Φ_1 and inner product $\langle\rangle_{H_{K_1}}$.
k_2 has its feature map Φ_2 and inner product $\langle\rangle_{H_{K_2}}$.

By linearity, we can have:

$\alpha k_1(x, z) = \langle \sqrt{\alpha}\, \Phi_1(x), \sqrt{\alpha}\, \Phi_1(z) \rangle_{H_{K_1}}$ and

$\beta k_2(x, z) = \langle \sqrt{\beta}\, \Phi_2(x), \sqrt{\beta}\, \Phi_2(z) \rangle_{H_{K_2}}$

Then:

$$k(x, z) = \alpha k_1(x, z) + \beta k_2(x, z)$$
$$= \langle \sqrt{\alpha}\, \Phi_1(x), \sqrt{\alpha}\, \Phi_1(z) \rangle_{H_{K_1}} + \langle \sqrt{\beta}\, \Phi_2(x), \sqrt{\beta}\, \Phi_2(z) \rangle_{H_{K_2}}$$
$$=: \langle [\sqrt{\alpha}\, \Phi_1(x), \sqrt{\beta}\, \Phi_2(x)], [\sqrt{\alpha}\, \Phi_1(z), \sqrt{\beta}\, \Phi_2(z)] \rangle_{H_{New}}$$

and that means $k(x, z)$ can be expressed as an inner product. Hence $k(x, z)$ is a kernel function. (Proved).

(b) $k(x, z) = k_1(x, z)\, k_2(x, z)$

For a kernel to be valid, it must correspond to a remapping of the input to a new feature space. That is $k(x, z) = \sum_i \Phi_i(x)\, \Phi_i(z)$ for some set of basis functions.

$$k_1(x, z)\, k_2(x, z) = \left(\sum_i \Phi_1(x_i)\, \Phi_1(z_i)\right) * \left(\sum_j \Phi_2(x_j)\, \Phi_2(z_j)\right)$$
$$= \sum_{i,j} \Phi_1(x_i)\, \Phi_2(x_j)\, \Phi_1(z_i)\, \Phi_2(z_j)$$

~~we can define Φ_K~~

let $\Phi_K(y) = \Phi_1(y_i)\, \Phi_2(y_j)$ [$\because$ each Φ function outputs a scalar]

Thus, we can finally write

$$k_1(x, z)\, k_2(x, z) = \sum_K \Phi_K(x)\, \Phi_K(z)$$

This shows that the product of two kernels creates a function with the same invariant that we started with. Hence, product of two kernels is also a kernel. (Proved)

c) $K(x,z) = h(K_1(x,z))$ where h is a polynomial function with positive co-eff.

Proof:

Since each polynomial term is a product of kernels with a positive co-efficient, the proof follows ~~from~~ by applying a) and b). Hence it is proved.

d) $K(x,z) = \exp(K_1(x,z))$

Proof

Since,
$$\exp(x) = \lim_{i \to \infty} \left(1 + x + \cdots + \frac{x^i}{i!}\right)$$

Now,
$$\exp(K_1(x,z)) \cancel{\lim_{i \to \infty} K_i(x,z)}$$
$$= \lim_{i \to \infty} K_i(x,z)$$

The proof follows from c).

e) $K(x,z) = \exp\left(\frac{-\|x-z\|_2^2}{\sigma^2}\right)$

Proof ~~we know the above as Gaussian~~

$$K(x,z) = \exp\left(\frac{-\|x-z\|_2^2}{\sigma^2}\right) = \exp\left(\frac{-\|x\|_2^2 - \|z\|_2^2 + 2x^Tz}{\sigma^2}\right)$$

$$\cancel{= \left[\exp\left(\frac{-\|x\|_2^2}{\sigma^2}\right)\exp\left(\frac{-\|z\|_2^2}{\sigma^2}\right)\right]\exp\left(\frac{2x^Tz}{\sigma^2}\right)}$$

$$\cancel{= [g(x)\,g(z)]\exp(K_1(x,z))}$$

$$= \exp\left(-\frac{x^Tx}{\sigma^2}\right)\exp\left(\frac{2x^Tz}{\sigma^2}\right)\exp\left(-\frac{z^Tz}{\sigma^2}\right)$$

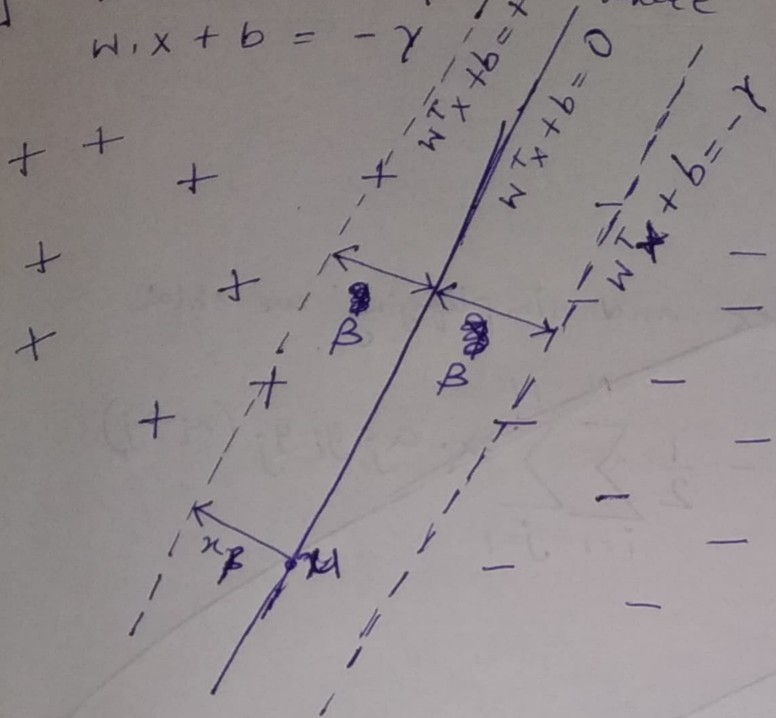
[We know the following kernels are valid, which are inferred from properties of kernels —
$K(x,z) = f(x)\,K_1(x,z)\,f(z)$ where $f(\cdot)$ is any function.]

Using the above property and (d), we can write together with the validity of the linear kernel. Also, product of two kernels is also a valid kernel. All these imply that the Gaussian Kernel legal.

1

$w.x + b = +\gamma$ where $\gamma > 0$ $\qquad y \in \{+1, -1\}$

$w.x + b = -\gamma$



~~margin = $\gamma / \|w\|$~~

$$w^T(x_1 + x_\beta) + b = \gamma$$

$$\underbrace{w^T x_1 + b}_{0} + w^T x_\beta = \gamma$$

$\Rightarrow w^T x_\beta = \gamma$

$\Rightarrow \|w\| \, \|x_\beta\| = \gamma$

$\Rightarrow \|w\| \beta = \gamma$

$\Rightarrow \boxed{\beta = \gamma / \|w\|}$ ← margin

Objective

$$\underset{w,\, b}{\text{maximize}} \quad \beta = \gamma / \|w\|$$

$$\text{s.t} \quad (w^T x_j + b)\, y_j \geq \gamma \quad \forall j$$

$\downarrow$

$$\underset{w,\, b}{\text{minimize}} \quad w^T w$$

$$\text{s.t} \quad (w^T x_j + b)\, y_j \geq \gamma \quad \forall j$$

So, we see that γ an arbitary constant does'nt affect the solution to the optimizing problem of finding maximum margin classifier.

2. From lecture slide 30, we have the SVM dual form as -

$$\max_{\alpha \geq 0} \min_{w, b} \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i [(w \cdot x_i + b) y_i - 1]$$

We can solve for optimal w, b as function of α :-

$$\frac{\partial L}{\partial w} = w - \sum_i \alpha_i y_i x_i \qquad \frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i$$

$$\Rightarrow 0 = w - \sum_i \alpha_i y_i x_i \qquad \Rightarrow 0 = \sum \alpha_i y_i$$

$$\Rightarrow w = \sum_i \alpha_i y_i x_i$$

Now,

we have to prove that given $\rho = \frac{1}{\|w\|}$

$$\frac{1}{\rho^2} = \|w\|^2 = \sum_i^N \alpha_i$$

let's take, $w = \sum_i \alpha_i y_i x_i$

Take

$$\|w\| = \pm \sum_i \alpha_i$$

[$\because x_i$ is a vector and $y_i \in \{1, -1\}$]

Squarring

$$\|w\|^2 = \sum_i \alpha_i$$ (Hence proved)