Problem 13

We begin with the naive solution, writing down our most explicit objective:

$$\text{Maximize } \gamma$$
$$\text{s.t. } \frac{y_i w^T x_i}{||w||} \geq \gamma$$

We observe that the scaling factor $||w||$ could instead be applied directly to γ: w

$$\text{Maximize } \frac{\gamma}{||w||}$$
$$\text{s.t. } y_i w^T x_i \geq \gamma$$

Which brings us nicely to the realization that we could remove γ altogether, replacing it with with a constant, and focus intead on maximizing only the scaling factor $\frac{1}{||w||}$, or conversely minimizing the norm:

$$\text{Minimize } ||w||^2$$
$$\text{s.t. } y_i w^T x_i \geq 1$$

We solve the Lagrangian duel, imposing lagrangian multipliers α for each point to constrain them on the correct side of the margins. The dual and the solution:

$$\text{Maximize} \sum_{i}^{n} \alpha_i - \frac{1}{2} \sum_{i,j}^{n} y_i y_j \alpha_i \alpha_j x_i^T x_j$$
$$\text{s.t. } \alpha_i \geq 0$$
$$\sum_{i}^{n} \alpha_i y_i = 0$$

Here we see the curious fact of the support vectors. The constraints in their lagrangian form (α) will only be active, naturally, for the data points which lie on the margin. Everything else will not need any constraints, and hence the α corresponding to that point will be zero. This is to say that our final $w*$ is defined by the following portion of our objective function:

$$\sum_{i,j \in S}^{n} y_i y_j \alpha_i \alpha_j x_i^T x_j$$

Where S is the set of points which lie on the margins. It is clear to see here that this is a linear product of scalar products of all the data points in S, hence, within the vector space spanned by those points.

Problem 14

Kernel Function

$$
\begin{aligned}
K(x,y) &= \langle \Phi(x), \Phi(y) \rangle \\
K(x,y) &= \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} x^n e^{-x^2/2} \frac{1}{\sqrt{n!}} y^n e^{-y^2/2} \\
K(x,y) &= e^{-x^2/2} e^{-y^2/2} \sum_{n=0}^{\infty} \frac{1}{n!} (xy)^n \\
K(x,y) &= e^{-x^2/2} e^{-y^2/2} e^{xy} \\
K(x,y) &= e^{xy - x^2/2 - y^2/2} \\
K(x,y) &= e^{\frac{1}{2}(x-y)(y-x)} \\
K(x,y) &= e^{-\frac{1}{2}(x-y)^2}
\end{aligned}
$$

Kernel in $\mathbb{R}^d$

Recognizing gaussianity when we see it, an easy choice is the multivariate flavor:

$$K(X,Y) = e^{-\frac{1}{2}(X-Y)^T(X-Y)}$$

Corresponding Feature Map

We begin by rewriting our Kernel function:

$$K(X,Y) = e^{X^T Y - ||X||^2/2 - ||Y||^2/2}$$

This allows us to more easily see the component parts:

$$\Phi(X) = \frac{1}{\sqrt{n!}} ||X||^n e^{-\frac{1}{2}||X||^2}$$

Problem 15

Product of Two Kernels

$$
\begin{aligned}
K_1K_2 &= \langle \Phi_1(x), \Phi_1(y) \rangle \langle \Phi_2(x), \Phi_2(y) \rangle \\
K_1K_2 &= \sum_{i=0}^{\infty} \Phi_1(x)_i \Phi_1(y)_i \sum_{j=0}^{\infty} \Phi_2(x)_j \Phi_2(y)_j \\
K_1K_2 &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Phi_1(x)_i \Phi_1(y)_i \Phi_2(x)_j \Phi_2(y)_j
\end{aligned}
$$

We can therefore define a new feature map:

$$
\Phi_3(x) = \Phi_i(x) \sum_{j=0}^{\infty} \Phi_j(x)
$$

And we have a scalar product:

$$
\begin{aligned}
K_1K_2 &= \sum_{i=0}^{\infty} \Phi_3(x)_i \Phi_3(y)_i \\
K_1K_2 &= \langle \Phi_3(x), \Phi_3(y) \rangle
\end{aligned}
$$

Sum of Two Kernels

$$
\begin{aligned}
K_1 + K_2 &= \langle \Phi_1(x), \Phi_1(y) \rangle + \langle \Phi_2(x), \Phi_2(y) \rangle \\
K_1 + K_2 &= \sum_{i=0}^{\infty} \Phi_1(x)_i \Phi_1(y)_i + \sum_{j=0}^{\infty} \Phi_2(x)_j \Phi_2(y)_j
\end{aligned}
$$

Here we see that this is simply the inner product of the two vectors concatenated together, so the corresponding new feature map can be defined as such, proving that this is indeed a valid kernel:

$$
\Phi_3(x) = \begin{bmatrix} \Phi_1(x) & \Phi_2(x) \end{bmatrix}
$$