SVM Quiz

1. Non-kernelised SVMs are linear classifiers

2. A unit vector is of magnitude or length 1

3. If **a** is a unit vector and **b** is any vector, then $a^T b$ is the length of **b** when projected to **a**

4. The decision boundary for a linear classifier in some feature space can be a straight line or a hyperplane. The essence of a linear decision boundary is that it is 'straight' -> In 1D a point, in 2D a line, in 3D a plane, in nD a (n-1)D hyperplane.

5. The decision boundary for a linear classifier is of the form $\underline{w}^T \underline{x} + w_0 = 0$ where **w** can be
   - a non-unit vector perpendicular to the decision boundary
   - a unit vector perpendicular to the decision boundary

6. In the SVM model formulation, the 'margin' is the distance from the decision boundary to the closest training point.

7. The is no difference between the hyperplane defined by $(\underline{w}, w_0)$ and the one defined by $(q\underline{w}, qw_0)$ where q > 1

8. The SVM model combines the kernelisation (a.k.a. the kernel trick) and maximum margin classification.

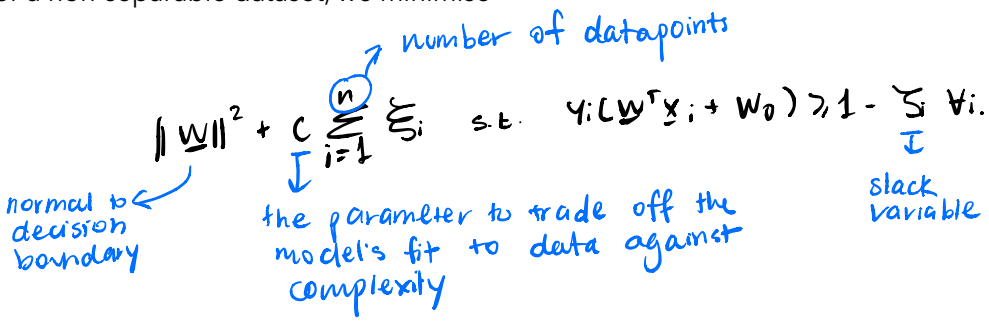
9. The max-margin optimisation problem, for training pairs $(\underline{x}_i, y_i)$ and model parameters $(\underline{w}, w_0)$ is:
   minimise $\|\underline{w}\|^2$ such that $y_i(\underline{w}^T \underline{x}_i + w_0) \geq 1 \quad \forall i$.

10. For a linear SVM, the form of the solution for the model parameters is: $\underline{w} = \sum_i \alpha_i y_i \underline{x}_i$

11. Important features of the solution in question 10 are:
    - most of the α_i are zero
    - if α_i is non-zero then $\underline{x}_i$ is called a support vector
    - if α_i is non-zero then $\underline{x}_i$ is on the margin
    - the optimisation problem to find the α_i has a unique global solution w/ no local minima.

12. If the dataset is not linearly separable, then we add slack variables >= 0 to all data points and minimise these along with the other model parameters.

13. For a non-separable dataset, we minimise



14. Assuming we have trained a linear SVM and obtained parameters , then to classify a new datapoint **x**, we

- predict class 1 if $\sum_{i=1}^{n} a_i y_i (\underline{x}_i^T \underline{x}) + w_0 \geq 0$, and class -1 otherwise.
- predict class 1 if sign $(\underline{w}^T \underline{x} + w_0)$ is positive, and class -1 if negative.
- predict class 1 if $\sum_{i=1}^{n} \alpha_i y_i (\underline{x}_i^T \underline{x}) + w_0 \geq 0$, and class -1 otherwise.

15. The kernelled SVM depends upon the fact that the classification of the new datapoint does not require us to know the actual datapoint values, but rather the dot product of the datapoint with the support vectors.

16. In a kernelled SVM, the kernel function takes two data points as parameters and computes the dot product of the two datapoint in the transformed space.

17. When we classify a new datapoint **x** with a kernelised SVM, we use the kernel function to compute the dot product of transformed **x** with each of the transformed support vectors, and then perform a linear weight sum of the results to see if it is positive or negative.

18. A kernel function *k* and a feature transformation ϕ correspond to one another if for two vectors x_1 and x_2
In the original space: $K(\underline{x}_1, \underline{x}_2) = \phi(\underline{x}_1^T \underline{x}_2)$

$\phi(x_1)^T \phi(x_2)$