Homework-2
Machine Learning 1

Instructions: All questions are mandatory. Do submit the answers in PDF format, with the file name being your name. The deadline for submission is December 16, 2022, at 11:59 p.m.

Problem 1. In real world applications, there are outliers in data. This can be dealt with using a soft margin, specified in a slightly different optimization problem as below (soft-margin SVM):

$$\min \frac{1}{2} w^T w + C \sum_i^N \xi_i \text{ where } \xi_i \geq 0$$
$$\text{s.t. } y^{(i)} \left(w^T x^{(i)} + b \right) \geq 1 - \xi_i$$

ξ_i represents the slack for each data point i, which allows misclassification of datapoints in the event that the data is not linearly seperable. SVM without the addition of slack terms is known as hard-margin SVM.
1. **[3 pt]** Intuitively, where do the data points lie relative to where the margin is when $\xi_i = 0$? Are all training data points classified correctly?
2. **[4 pt]** Intuitively, where does each data point lie relative to where the margin is when $0 < \xi_i \leq 1$? Are all training data point classified correctly?
3. **[3 pt]** Intuitively, where does each data point lie relative to where the margin is when $\xi_i > 1$? Are all training data points classified correctly?

Problem 2. Support Vector Machines can be used to perform non-linear classification with a kernel trick. Recall the hard-margin SVM from class:

$$\min \frac{1}{2} w^T w$$
$$\text{s.t. } y^{(i)} \left(w^T x^{(i)} + b \right) \geq 1$$

The dual of this primal problem can be specified as a procedure to learn the following linear classifier:

$$f(x) = \sum_i^N \alpha_i y_i \left(x_i^T x \right) + b$$

Note that now we can replace $x_i^T x$ with a kernel $k\left(x_i, x\right)$, and have a non-linear decision boundary.

In Figure 5, there are different SVMs with different shapes/patterns of decision boundaries. The training data is labeled as $y_i \in \{-1, 1\}$, represented as the shape of circles and squares respectively. Support vectors are drawn in solid circles/squares. Match the scenarios described below to one of the 6 plots (note that one of the plots does not match to anything). Each scenario should be matched to a unique plot. Explain in less than two sentences why it is the case for each scenario.

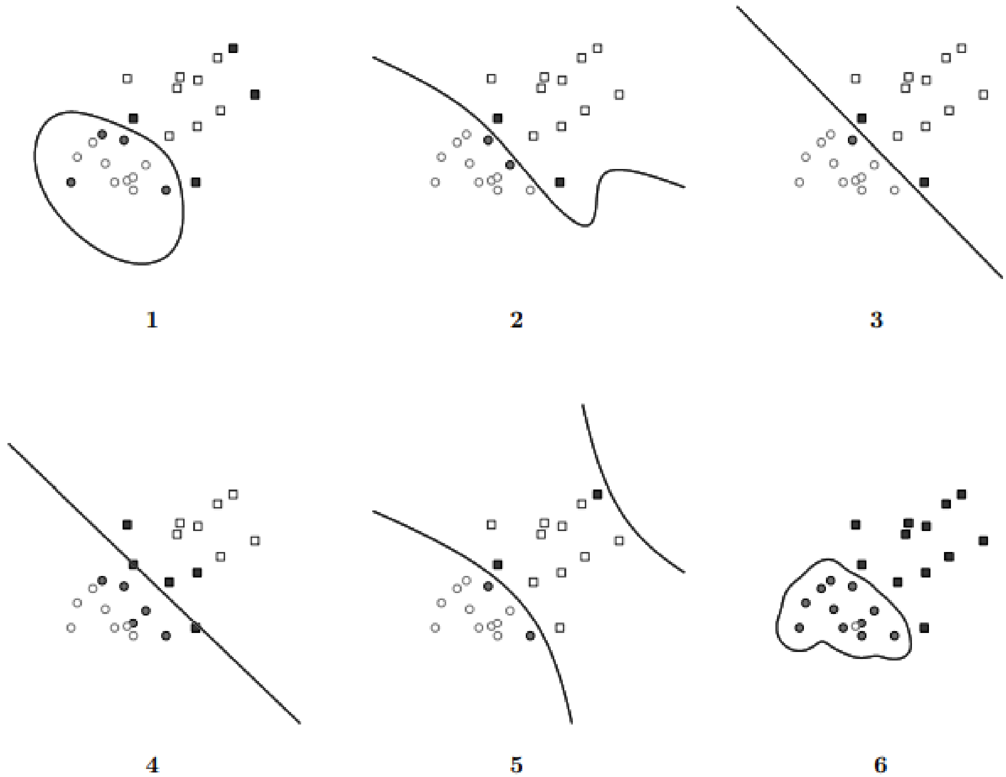

Figure 1: SVM boundaries

1. **[2 pt]** A soft-margin linear SVM with $C = 0.02$.
2. **[2 pt]** A soft-margin linear SVM with $C = 20$.
3. **[2 pt]** A hard-margin kernel SVM with $k(u, v) = u \cdot v + (u \cdot v)^2$
4. **[2 pt]** A hard-margin kernel SVM with $k(u, v) = \exp\left(-5\|u - v\|^2\right)$
5. **[2 pt]** A hard-margin kernel SVM with $k(u, v) = \exp\left(-\frac{1}{5}\|u - v\|^2\right)$

Problem 3. Suppose we have the following data on seven variables $x_1, \cdots, x_7$ and the output y, given as follows:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y
0	0.94	0.37	0.76	0.56	0.77	0.51	0.80	0.66
1	-0.88	-0.33	-0.68	-0.51	-0.75	-0.47	-0.76	-0.60
2	1.32	-0.31	0.78	0.23	0.47	0.10	0.90	0.50
3	-0.21	0.79	0.13	0.46	0.27	0.54	0.05	0.29
4	0.75	-0.70	0.28	-0.22	-0.07	-0.34	0.39	0.03
5	-0.33	-1.30	-0.66	-0.98	-0.86	-1.06	-0.57	-0.82
6	1.27	0.81	1.14	0.96	1.08	0.92	1.15	1.04
7	-0.60	0.84	-0.13	0.36	0.35	0.48	-0.24	0.12
8	0.15	-1.35	-0.35	-0.85	-0.47	-0.98	-0.22	-0.60
9	-0.33	0.86	0.07	0.46	0.37	0.56	-0.03	0.26

Note: For any of the problems below, do not use the constant term in the regression.
(i) **[5 pt]** Find the linear regression to this data using the closed form expressions (can use calculators). Does the formula work? If not, explain why not.
(ii) **[5 pt]** Fit a linear ridge regression model to this data using the closed form expressions (can use calculators).
(iii) **[5 pt]** Fit a linear lasso regression model to this data using a computer program - can use packages.
(iv) **[12 pt = 3+3+3+3]** Use the following subset selection methods to choose the two features that best explain the data:
(a) Forward Stepwise (b) Backward Stepwise (c) Forward Stagewise (d) Best Subset (All the $\binom{7}{2}$ subsets)
(v) **[3 pt]** Explain all your observations based on the results.