

Support Vector Machines (SVM)

Support Vector Machines (SVM)

Summary

These exercises aim a deeper understanding of how a support vector machine works. The firts questions are more conceptual. In the last two exercises you will get to use a library to train a SVM and investigate some of its properties.

Exercise 1

This exercise is about hyperplanes in two dimensions. Imagine we have a 2-dimensional feature space with features X_1 and X_2 .

- Sketch the hyperplane $1 + 3X_1 - X_2 = 0$. Indicate the set of points for which $1 + 3X_1 - X_2 > 0$, as well as the set of points for which $1 + 3X_1 - X_2 < 0$.
- On the same plot, sketch the hyperplane $-2 + X_1 + 2X_2 = 0$. Indicate the set of points for which $-2 + X_1 + 2X_2 > 0$, as well as the set of points for which $-2 + X_1 + 2X_2 < 0$.

Exercise 2

Here we explore hard-margin SVM (maximal margin classifier) on a toy data. We are given $n = 7$ observations in $p = 2$ dimensions. For each observation, there is an associated class label.

Obs.	X_1	X_2	Y
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue

- Sketch the observations
- Sketch a separating hyperplane.
- Sketch the optimal separating hyperplane, and provide the equation for this hyperplane.
- Describe the classification rule, something along the lines of "Classify to Red if $\beta_0 + \beta_1 X_1 + \beta_2 X_2 > 0$, and classify to Blue otherwise".
- On your sketch, indicate the margin for the maximal margin hyperplane.
- What are the support vectors for your classifier?
- Argue that a slight movement of the seventh observation would not affect the maximal margin hyperplane.
- Sketch a hyperplane that is not the optimal separating hyperplane, and provide the equation for this hyperplane.
- Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane.

Exercise 3

Explain in your words:

- What is the fundamental idea behind Support Vector Machines?
- What is a support vector?
- Is there a difference between support vectors in hard-margin and soft-margin?
- What is the general role of the regularization parameter C in soft-margin SVM?

Exercise 4

We have seen that in $p = 2$ dimensions, a linear decision boundary takes the form $\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$. We now investigate a non-linear decision boundary.

- Sketch the curve $(1 + X_1)^2 + (2 - X_2)^2 = 4$.
- On your sketch, indicate the set of points for which $(1 + X_1)^2 + (2 - X_2)^2 > 4$, as well as the set of points for which $(1 + X_1)^2 + (2 - X_2)^2 \geq 4$
- Suppose that a classifier assigns an observation to the blue class if $(1 + X_1)^2 + (2 - X_2)^2 > 4$. and to the red class otherwise. To what class is the observation $(0, 0)$ classified? $(-1, 1)$? $(2, 2)$? $(3, 8)$?
- Argue that while the decision boundary is not linear in terms of X_1 and X_2 , it is linear in terms of X_1, X_1^2, X_2 , and X_2^2 .

Using Scikit-Learn SVM

Exercise 5

In this exercise, we use svm.SVC from sklearn library with different kernels and different values for regularizationn hyperparameter C that penalizes the training samples for being on the wrong side of the margins.

Try the following with **two** different synthetic datasets (e.g., make_blobs and make_circles from sklearn.datasets)

- Always start with normalysing your data.
- Classify your data with sklearn.svm.SVC function:
 - use a linear kernel.
 - try differet values for hyperparameter C and compare the results.
- Now use a polynomial kernel on your data, try different valued for C, and compare the results.

Answer this question: Imagine you have trained an SVM classifier, but it seems to underfit the training set. Should you increase or decrease C?

Now try an RBF kernel on your data. The RBF kernel is defined as $K(\mathbf{x}, \mathbf{z}) = e^{-\frac{\|\mathbf{x}-\mathbf{z}\|^2}{2\sigma^2}}$. It can be also formulates as $K(\mathbf{x}, \mathbf{z}) = e^{-\gamma\|\mathbf{x}-\mathbf{z}\|^2}$ where $\gamma = \frac{1}{2\sigma^2}$.

- try differet values for hyperparameter C and γ (gamma) and compare the results.

Answer this question: If you have trained an SVM classifier with an RBF kernel, but it seems to underfit the training set. Should you increase or decrease γ (gamma)? What about C?

Exercise 6

Now use the datasets Ex1-training.csv and Ex1-test.csv and Classify them with sklearn.svm.SVC function.

- Compare the decision boundaries with what you have obtained from classification methods in previous weeks, e.g, Decision Trees, knn, LR, LDA and QDA.