

Exam questions:

Chapter 9:

- What is the maximum margin classifier? Explain how you classify a new observation?
- What is a support vector in the context of a maximum margin classifier? Do you think that the support vectors are representative for the maximum margin hyperplane in the test data set when p is large?
- Support vector classifiers are able to handle the cases when classes are not separable. Yet in many cases, even if a separating hyperplane does exist, then there are instances in which a classifier based on a separating hyperplane is not desirable. Explain the situation!
- Explain how the variable C in the optimization problem controls the bias-variance trade-off:

$$\begin{aligned} & \underset{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n, M}{\text{maximize}} && M \\ & \text{subject to} && \sum_{j=1}^p \beta_j^2 = 1, \\ & && y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i), \\ & && \epsilon_i \geq 0, \quad \sum_{i=1}^n \epsilon_i \leq C, \end{aligned}$$

How does C influence the number of support vectors?

What is the function of variable M ?

- Why is the constrain on the beta variables needed?

$$\begin{aligned} & \underset{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n, M}{\text{maximize}} && M \\ & \text{subject to} && \sum_{j=1}^p \beta_j^2 = 1, \\ & && y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i), \\ & && \epsilon_i \geq 0, \quad \sum_{i=1}^n \epsilon_i \leq C, \end{aligned}$$

- Explain the purpose of slack variables? Write down the equation of the maximum margin classifier and explain the relation between the slack variable and the objective of maximizing the margin. Explain using a figure!
- Why is a support vector classifier robust to the behavior of observations far from the hyperplane. Is this the same in LDA? Explain!
- What is a support vector machine? How does it differ from a SV classifier?
- How do you construct an ROC-curve for a SVM?
- What is the difference between classification and prediction in the binary responds setting?
- Consider the case with ordinal independent variables and a continuous dependent variable. What are the pros and cons of the ANOVA approach and the regression approach?

- How can you extend the principle of SVM to multiple classes? Draw some examples! Is there a similarity with LDA?