Machine Learning and Bayesian Inference

Problem sheet II: support vector machines and general methodology

Sean B. Holden

1 Support vector machines

1. **Slide 16** provides an alternative formulation of the maximum margin classifier based on maximizing γ directly with suitable constraints.

   Apply the KKT conditions to this version of the problem. What do they tell you about the solution, and how does it differ from the version developed in the lectures?

2. **Slide 27** states the dual optimization problem for the maximum margin classifier. Provide a full derivation.

3. **Slide 30** states the optimization problem that needs to be solved to train a support vector machine

$$\underset{\mathbf{w},w_0,\boldsymbol{\xi}}{\text{argmin}} \frac{1}{2}||\mathbf{w}||^2 + C\sum_i \xi_i \text{ such that } y_i f_{\mathbf{w},w_0}(\mathbf{x}_i) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \text{ for } i = 1,\ldots,m.$$

   Apply the KKT conditions to this version of the problem. What do they tell you about the solution?

2 Machine learning methods

1. **Slide 57** uses the following estimate for the variance of a random variable:

$$\sigma^2 \simeq \hat{\sigma}^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i - \hat{X}_n)^2\right].$$

   Show that this estimate is unbiased; that is,

$$\mathbb{E}\left[\hat{\sigma}^2\right] = \sigma^2.$$

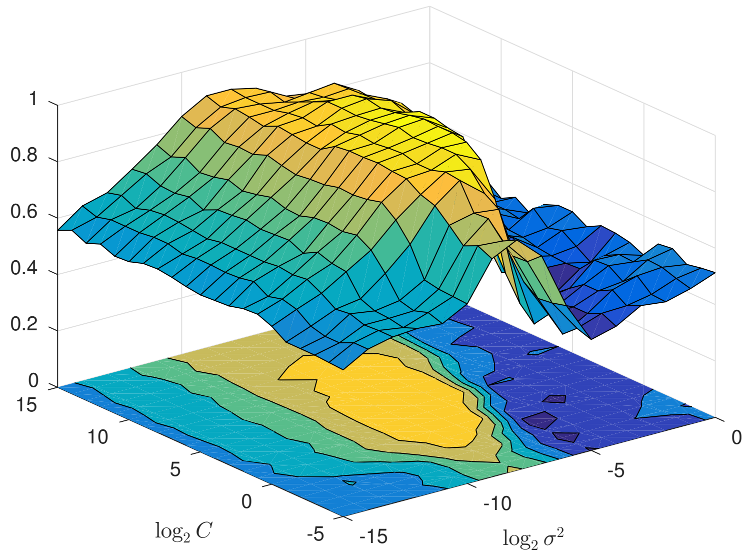
2. Show that if a random variable has zero mean then dividing it by its standard deviation σ results in a new random variable having zero mean and variance 1. Show that in general multiplying a random variable having mean μ and variance σ^2 by $\sqrt{c}$ alters its mean to $\sqrt{c}\mu$ and its variance to $c\sigma^2$.

3. Verify the expression in point 4 on **slide 60**.

3 Making it all work

Probably the best way to get a feel for this material is to write some code that implements it. In particular, can you reproduce something like the hyperparameter search graph?

Using crossvalidation to optimize the hyperparameters C and σ^2.



In order to do this I don't suggest you attempt to implement SVMs from scratch—having said that, if you can find a suitable, general constrained optimization library it's not too hard. A quicker approach initially is to find a good SVM library in a system such as Matlab or R. You will need to generate the spiral data set and implement a search using cross-validation to assess possible hyperparameter values.

4 Old exam questions

Most of the material here is quite new, so the only relevant past question is:

- 2016, paper 8, question 2.