EPFL, Autumn 2024
Yoav Zemel
Guillaume Obozinski

Ho Yun
Shivang Sachar

Statistical Machine Learning

Exercise sheet 7

Practical exercise

Solution: See uploaded code for solutions to the implementation exercises.

Exercise 7.1 (Nadaraya-Watson and LOO CV) In this exercise we consider using the Nadaraya-Watson (NW) estimator for regression.

(a) Express the Nadaraya-Watson estimator as $\widehat{\boldsymbol{y}} = \mathbf{S}\boldsymbol{y}$, where $\mathbf{S}$ is an $n \times n$ matrix whose values only depend on the inputs $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ and you should write down explicitly.

(b) Generate n simulated data $\{(y_1, x_1), \ldots, (y_n, x_n)\}$ based on the relationship

$$y = x^2 \cos(x) + \epsilon,$$

where x is a normally distributed random variable with mean 0 and variance 4 and ϵ is a normally distributed random variable with mean 0 and variance 0.25.

(c) Code up the NW estimator in a function, fit it on the simulated data, plot the data on a grid of $x's$ from -10 to 10, and experiment with different values of bandwidth. Here is a template to get you started:

```
nw <- function(x, X, Y, h, K) {

        # Arguments
        # x: evaluation points
        # X: vector (size n) with the predictors
        # Y: vector (size n) with the response variable
        # h: bandwidth
        # K: kernel

    # << Insert code here >>
}
```

(d) What is $\widehat{f}^{-i}$ for the NW estimator? Code this up and verify with part (a) that indeed,

$$y_i - \widehat{f}^{-i}(\boldsymbol{x}_i) = \frac{y_i - \widehat{f}(\boldsymbol{x}_i)}{1 - \mathbf{S}_{ii}}.$$

(e) How would you propose to choose the bandwidth h for this estimation problem?

Solution: Look at the CV score. either explicitly or looking at diagonals of the matrix S to get the leave one out CV score explicitly.

(f) With your chosen bandwidth h, plot your predictions and verify graphically that your chosen bandwidth is reasonable.

Exercise 7.2 (Nadaraya-Watson, ROC, Precision, Recall) In this exercise we consider using the NW estimator for classification. Load the R script `R solution template.R` to get you started. If you have any questions (especially coding related), please do not hesitate to ask during the session.

(a) Explain how you can adapt your estimator in Exercise 7.2 to perform binary classification.

Solution: Here, we can use the NW estimator to estimate the target function $\mathrm{E}(Y|X)$ and define the 'plug-in' estimator $\widehat{f}$ such that $\widehat{f}(x) = 1$ if $\sum_i w_i(x) y_i > 1/2$

(b) Based on the data you simulated in Exercise 7.2(b), generate n simulated data pairs $\{(x_1, z_1), \ldots, (x_n, z_n)\}$ where x_i is simulated as before and $z_i = 1/(1 + \exp(-y_i)) > 0.5 \in \{0, 1\}$, $i = 1, \ldots, n$.

(c) Set h=0.2 in your NW estimator. Calculate the confusion matrix for your classifier.

(d) Calculate the misclassification rate, precision and recall of your classifier.

(e) Plot the ROC curve and calculate the AUC of your classifier using the `auc` function from the `pROC` library.

(f) Suppose that you want to your classifier to achieve a false positive rate of 20%, what should you do?

(g) Suppose now that the cost of a false positive is 4 times that of a false negative. Can you think of a classification setting where this is the case? Does it still make sense to use the misclassification rate (0-1 loss) for your cost function? Suggest a sensible cost function and suggest how you could adapt your classifier to achieve the minimum cost.

Solution: Change the threshold for the 'plug-in' estimator.

(h) *BONUS question:* In practice, we also have to vary the bandwidth. Suggest how you would do this given your new loss function given in part (g). *Hint: Look back at Exercise 7.2(f)!*