

STATS 415 Homework 10

Due Thursday April 5, 2018

Please include your name, username, and lab section (number or time or GSI). A point will be taken off homework without the section info. Turn in a printout of your homework in the lecture or in your GSI's mailbox across room 305A West Hall, no later than 5pm on the due date.

1. Textbook Section 9.7, p. 368, Conceptual exercise 3.
2. This question uses the same `crabs` data used in Homework 9. Use the following code to split the data into training and test sets:

```
set.seed(45678)
blueMale = which(sp == "B" & sex == "M")
orangeMale = which(sp == "O" & sex == "M")
blueFemale = which(sp == "B" & sex == "F")
orangeFemale = which(sp == "O" & sex == "F")
train_id = c(sample(blueMale, size = trunc(0.80 * length(blueMale))),
sample(orangeMale, size = trunc(0.80 * length(orangeMale))),
sample(blueFemale, size = trunc(0.80 * length(blueFemale))),
sample(orangeFemale, size = trunc(0.80 * length(orangeFemale))))
crabs_train = crabs[train_id, ]
crabs_test = crabs[-train_id,]
```

- (a) Fit a linear support vector machine to the data with various values of `cost`, in order to predict Species from the **five numerical measurements**. **Omit the variable Sex** for this homework. Report the cross-validation errors associated with different values of `cost`. Comment on your results and make some relevant plots.
- (b) Fit nonlinear SVMs with radial and polynomial kernels, with different values of `gamma` and `degree` and `cost`. Report the cross-validation errors associated with different values of `cost`. Comment on your results and make some relevant plots.

Please limit your solution to Problem 2 to at most 4 pages.