

wbs

WARWICK BUSINESS SCHOOL
THE UNIVERSITY OF WARWICK

For the Change Makers

Dr. Iman Ahmadi
Assistant Prof. of Marketing

**Marketing & Strategy
Analytics:
Supervised Learning: K-NN**

Exercise 7.2 – Predicting Cancer (I/II)

- You are asked to come up with a model that helps to predict whether a patient has a breast cancer or not.
- Use dataset "Cancer" and build a model in R using knn command (dependent variable: 'diagnosis')
 - use the first 469 observations as your train dataset
 - use the rest observations (i.e., the last 100 observations) for evaluating your model

("Cancer" R code)



Exercise 7.2 – Predicting Cancer (II/II)

- **Dataset includes:**

- 569 observations
- 31 variables
 - Independent variables are a list of 30 laboratory measures:
 - radios
 - texture
 - area
 - dimension
 - and ...
 - Dependent variable: diagnosis

- **The winner has the lowest error rate in the test dataset**

(Hint: increase/decrease k and see if you see decrease in your error rate)



Thank You!

Iman.Ahmadi@wbs.ac.uk

Room No.: 3.207