COSC522 – Homework #1 (Due 09/05)

Problem 1 (65): Suppose you are doing a test on covid diagnosis using only temperature. If the patient does not have covid, he/she is labeled class 1; otherwise, class 2. Hence this is a 1-D, 2-class machine learning problem for the classification of covid patient with one feature. Assume the density function of each class can be adequately represented by univariate Gaussians, with μ_1=95, σ_1=4, μ_2=103, σ_2=2.

(1) (15) Sketch the two density functions on the same figure using pencil and paper (i.e., without Python or MATLAB or other software package). Assume equal prior probability, predict how many decision regions there would be. (Grading: pay attention to the relative position, relative height, relative width of the Gaussians)

(2) (35) Assume equal prior probability,

  a. (5) If x=100, which class does x belong to? Use the MAP (or MPP) method. Show detailed steps.

  b. (15) Find the exact decision boundary using analytical methods instead of the sketch.

  c. (15) Solve for the overall probability of error. Provide details.

(3) (15) Assume that P(ω_1)=0.8, P(ω_2)=0.2.

  a. (10) Use a software package to draw the posterior probability.

  b. (5) Redo question (2)a.

Problem 2 (25 pts): Instead of modeling the density function using Gaussian, assume here we use uniform distribution. Let class 1 be the actual negative cases with a uniform pdf on the interval [95, 101]. Let class 2 be the actual positive case with a uniform pdf on the interval [98, 103]. Assume equal prior probability. Given a decision boundary of 99, that is, any temperature reading less than 99 indicates covid negative and any temperature reading above 99 indicates covid positive,

(1) (5) Plot (or sketch, whichever is easier) the pdfs on the same figure.

(2) (5) What is the probability for false-negative? That is, the patient actually has covid but the classifier says otherwise.

(3) (5) What is the probability for false-positive? That is, the patient is negative but our classifier says otherwise.

(4) (5) Is 99 the optimal decision boundary in Baysian sense? If not, what is the optimal decision boundary that minimizes the overall probability of error?

(5) (5) Can you do better than the optimal (or minimum) probability of error by adjusting the prior probability?

Problem 3 (10 pts): Read Breiman's paper on statistical model and write an essay of [1,500, 2,500] characters (including space).