Exam Pattern Recognition
Wednesday, February 2, 2022
Short Answers to Q 1-3

Question 1: Linear Regression

(a) Of course there are different ways to phrase this. One example is: "If BMI increases with one unit, then the number of COVID-19 related deaths per million inhabitants is expected to increase with about 28."

(b) The coefficient of `BMI` is significant at $\alpha = 0.05$. The relevant p-value `Pr(>|t|)` is 0.0125 which is smaller than 0.05.

(c) About 10.6%, because the reported R^2 is 0.1063.

(d) The reported average male BMI of Ukraine is 17.09, so:

$$-581.94 + 28.08 \times 17.09 \approx -102$$

Of course a negative number of deaths doesn't make any sense. In this case the negative number is probably due to an error in the reported BMI of Ukraine. The value 17.09 is an outlier in the list of BMI values.

(e) Due to the interaction term, the effect of BMI on the number of deaths now depends on the value of life expectancy. This is most clear if we rearrange the equation a bit:

```
Deaths = 9154.280-408.511*BMI-125.768*LIFE + 5.635*BMI*LIFE =
9154.280+(5.635*LIFE - 408.511)*BMI - 125.768*LIFE
```

For example, if `LIFE=80`, then a one unit increase in `BMI` will lead to an *increase* in the number of deaths of about 42, but if `LIFE=70` it will lead to a *decrease* in the number of deaths of about 14.

Question 2: Logistic Regression

(a) The higher the value of `ink`, the higher the probability that the digit written is a five. That makes sense, since writing a five tends to consume more ink than writing a one.

(b) In an all white image both features have the value 0, so only the intercept plays a role:

$$\frac{1}{1+e^{-4.73567}} \approx 0.99$$

So the estimated probability that the digit is a one is about 99%.

(c) Rounding the coefficients to two decimal places:
If $0.16 \times \texttt{ink} + 0.12 \times \texttt{asymmetry} > 4.74$ then predict five, otherwise predict one.

(d) You can eyeball from the plot that the densities intersect at about `ink` = 20. To the left of that value, the probability density of `ink` for digit one is higher than the probability density of `ink` for digit five. To the right of 20 it is the other way around. To minimize classification errors we assign to the class with the highest probability, so the rule becomes: If `ink` > 20 predict five, otherwise predict one.

Question 3: Support Vector Machines

(a) (3,6) and (5,8) are the two points from opposite class that are closest to each other, and their perpendicular bisector is given by

$$x_1 + x_2 - 11 = 0$$

It is not true, as a general rule, that the perpendicular bisector of the two closest points is the maximum margin decision boundary. However, in case all other points are further removed from this line than the two closest points themselves, then it is. We can see that this is the case for the given data set. We still have to scale the coefficients, so that for the points $\mathbf{x}_i$ closest to the decision boundary we have $y(\mathbf{x}_i)t_i = 1$. The closest points are (3,6) and (5,8). Let's take (5,8) to compute its current y value:

$$5 + 8 - 11 = 2$$

That's twice as big as it should be, so we divide all coefficients by 2 to get the properly scaled decision boundary:

$$0.5x_1 + 0.5x_2 - 5.5 = 0$$

(b) (3,6) and (5,8)

(c) We have $y(x_1, x_2) = 0.5x_1 + 0.5x_2 - 5.5$. Filling in $x_1 = 5$ and $x_2 = 7$ gives:

$$0.5 \times 5 + 0.5 \times 7 - 5.5 = 6 - 5.5 = 0.5$$

Since the value is positive, we predict the class +1.

(d) All data points except (3,6) and (5,8) (the support vectors) have Lagrange multiplier 0. Let's call the Lagrange multiplier of (3,6) a_1, and of (5,8) a_2. From equation (2) it follows that $a_1 = a_2$, so for simplicity, let's call this common value a. From equation (1) it follows that:

$$\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = a \begin{bmatrix} 5 \\ 8 \end{bmatrix} - a \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

Therefore, $a = 0.25$. Hence, finally, both support vectors have a Lagrange multiplier of 0.25.