Machine learning for data science I

7 June 2022

Surname, name (all caps) ______________________________

Student ID: ______________________________

This is a closed book exam.

Write clearly and justify your answers.

Time limit: 90 min.

Question:	1	2	3	4	5	Total
Points:	20	20	20	20	20	100
Score:						

[20] 1. We are dealing with a binary classification problem in a high-dimensional space. However, the data points are known to lie on a lower-dimensional hyperplane. An example of such data are images with a vector description by pixels. How well would certain models perform in such high-dimensional space compared to learning directly on a low-dimensional space, where the data points are known to lie on? Explain your answers.

1. k-nearest neighbors with Euclidean distance
2. decision tree
3. logistic regression
4. support-vector machine

[20] 2. Answer the following questions about the intercept of a linear regression model.

1. Is the intercept zero if the response variable is zero-centered (its mean is equal to zero)? Explain your answer.
2. What is the value of the intercept if the feature variables are zero-centered but the response variable is not? Prove your result.

[20] 3. In the target variable scenarios below, which distribution would you choose for the likelihood if you were doing parametric modeling for the purpose of prediction and generation of new data. Justify your answer:

1. Student grades at FRI.
2. Number of calls to customer support.
3. SO2 pollution levels.
4. A student's choice of undergraduate study program.
5. Air temperature.
6. Free throw shooting % in basketball.
7. Position where the dart hit the round dartboard. The dart can also miss completely - all misses are treated the same, regardless of how much you miss.

[20] 4. Kernelize the Perceptron algorithm described with the following pseudocode.

```
Initialize w⃗ = 0⃗ ;
m = 1;
while m > 0 do
    m = 0 ;
    for (x⃗_i, y_i) ∈ D do
        if y_i(w⃗^T x⃗_i) ≤ 0 then
            w⃗ = w⃗ + y_i x⃗_i;
            m = m + 1;
        end
    end
end
```

[20] 5. The inverse-Gamma distribution is a continuous distribution on positive reals with two positive parameters α and β and the following PDF:

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}}.$$

Let y_i, $i = 1..n$ be our data. We assume an iid normal likelihood with known μ (it is a known constant) and σ^2 as the only parameter. Show that if the inverse-Gamma distribution is chosen as a prior distribution on σ^2 (not σ!), then the posterior $p(\sigma^2|\mu, y)$ is also inverse-Gamma. That is, inverse-Gamma is conjugate for the normal likelihood variance when the mean is constant. What are the parameters of the posterior?