CPSC 383 – Winter 2025
Tutorial worksheet: More ML models

Your TA will start with a short introduction explaining a bit about the models and how they work. After this, play around with the tools provided for each model to get a sense for what is going on. For this part, work and discuss your ideas in groups of 3–4 students. Once you think you all have a good understanding, answer the listed questions. To receive credit for this worksheet, you will need to explain your answers to your TA during the tutorial and have them approved.

Part 1: SVM

Tool: `https://greitemann.dev/svm-demo`

Notes:

- Stick with a simple linear kernel, as kernel transformations are beyond the scope of this course.
- To start, keep the hyperparameter ν as low as possible. You'll adjust it a bit later in the final questions.

1. What type of problem is the SVM solving? It is supervised or unsupervised? If it is supervised, is this a classification or regression problem?
2. What is the format of the input data? What about output data (predictions)?
3. What model class is the SVM searching in? What are the model parameters?
4. Consider the "road" that the SVM finds between the two classes of data points in the case where they are linearly separable. What happens if you add an extra data point that is:
   (a) On the correct side of the road, outside the boundary of the road?
   (b) On the correct side of the road, within the boundary of the road?
   (c) On the wrong side of the road?
5. Can you create a data set where the SVM algorithm leads to overfitting (keeping ν as small as possible)?
6. Now, start adjusting ν on your above example. What effect does this have? Does it give a better or a worse model of the data, and why?

Part 2: Polynomial regression

Tool: `https://visualize-it.github.io/polynomial_regression/simulation.html`

1. What type of problem is the polynomial regression algorithm solving? Is it supervised or unsupervised? If it is supervised, is it a classification or regression problem?
2. What is the format of the input data? What about output data (predictions)?
3. What model class is the regression algorithm searching in? What are the model parameters?
4. Can you give an example of a dataset where **none** of the polynomials achieve a good fit?
5. Can you create a data set where the regression algorithm leads to overfitting?
6. How can the choice of k be adjusted to avoid overfitting? Are there any reasons you might not want to do this?