Model Zoo 3

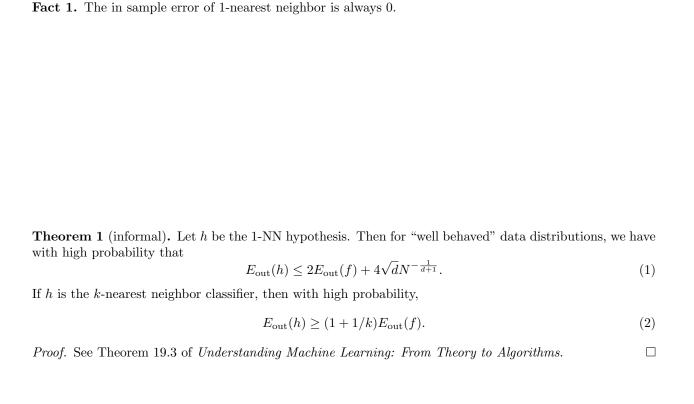
The textbook has detailed explanations of all of the models we've covered in the model zoo in the online supplements located at

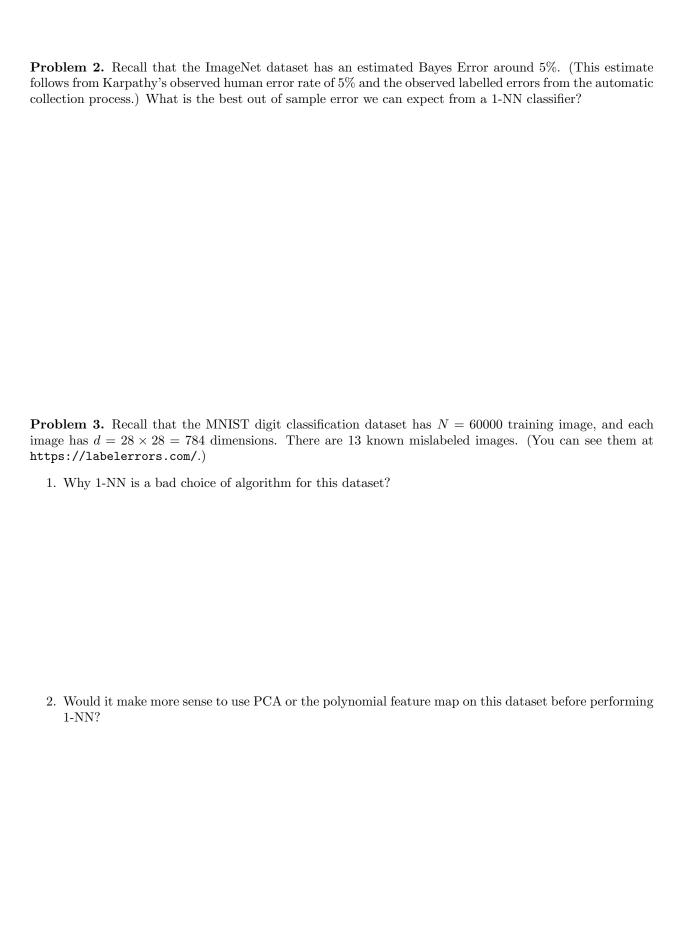
https://amlbook.com/eChapters.html

The online supplements are all password protected with password Paraskavedekatriaphobia. For the most part, these online supplements are technical and provide relatively little insight into practical applications. The section on nearest neighbor algorithms (e-Chapter 6), however, is relatively readable.

Nearest Neighbor Methods

Problem 1. Describe the k-nearest neighbor classifier.





Support Vector Machines (SVMs)

There are three standard and equivalent interpretations of SVMs:

- 1. The hypothesis class $\mathcal{H}_{perceptron}$ with the hinge loss $\ell(\mathbf{x}; \mathbf{w}) = \max(1 \mathbf{w}^T \mathbf{x}, 0)$.
- 2. The large margin classifier.
- 3. A smooth generalization of the k-nearest neighbor algorithm.

Note 1. The following theorem is copied directly from Theorem 8.5 in Chapter 8 of the *Learning from Data* textbook. Most of this chapter is more technical than needed for this course, but the explanation of the theorem is straightforward and worth reading.

Theorem 2. Suppose the input space is the ball of radius R in \mathbb{R}^d . That is, $\|\mathbf{x}\|_2 \leq R$. Then,

$$d_{\rm VC} \le \lceil R^2/\rho^2 \rceil + 1. \tag{3}$$

Problem 4. Describe the dual learning problem and the kernel trick.

Common sample kernel functions include:

kernel name	$K(\mathbf{x}_1,\mathbf{x}_2)$	feature dimensions (\tilde{d})
linear	$\mathbf{x}_1^T\mathbf{x}_2$	d
polynomial	$(\gamma \mathbf{x}_1^T \mathbf{x}_2 + r)^Q$	$\Theta(d^Q)$
gaussian	$\exp(-\gamma \ \mathbf{x}_1 - \mathbf{x}_2\ _2^2)$	∞
sigmoid	$\tanh(\gamma \mathbf{x}_1^T \mathbf{x}_2 + r)$	∞