STAT3007/7007 Deep Learning, Tutorial 2

2022 Semester 2

1. (Ridge regression) We examine some claims about ridge regression in the lecture slides.

(a) (Lecture 2 slide 21) We see that ridge regression (without the bias term) minimizes the regularized SSE

$$\beta_n = \operatorname*{argmin}_{\beta \in \mathbf{R}^d} \Big(\sum_{i=1}^{n} (\mathbf{x}_i^T \beta - y_i)^2 + \lambda ||\beta||_2^2 \Big).$$

Show that $\beta_n \in \mathbf{R}^d$ is also the solution to OLS (without the bias term) with training examples $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n), (\sqrt{\lambda}\mathbf{e}_1, 0), \ldots, (\sqrt{\lambda}\mathbf{e}_d, 0)$, where $\mathbf{e}_i$ is the i-th one-hot vector (i.e., the vector with the i-th entry being 1 and all others being 0).

(b) (Lecture 2 slide 22) Show that $\lambda I + \mathbf{x}^\top \mathbf{x}$ is invertible for any $\lambda > 0$ and any matrix $\mathbf{x} \in \mathbf{R}^{n \times d}$.

2. (Naive Bayes classifier) Prove the following claim (Lecture 3 slide 15): the naive Bayes classifier maximizing the likelihood $\prod_{i=1}^{n} p(\mathbf{x}_i, y_i)$ is given by

$$\hat{p}(y) = n_y / n,$$
$$\hat{p}(x_i \mid y) = n_{y,x_i} / n_y,$$

where n_y is the number of times class y appears in the training set, and n_{y,x_i} is the number of times attribute i takes value x_i when the class label is y.

Hint: (a) This is a constrained optimization problem. (b) Maximize the log-likelihood instead. (c) Use the method of Lagrangian multipliers.