CS680, Spring 2020, Assignment 3

Zhijie Wang, [redacted]

May 31, 2020

Exercise 1

1. Attached as `ex1_1.py`.
2. Attached as `ex1_2.py`. Estimated Bayes error is around 0.38 when n is large enough (> 100000).
3. Attached as `ex1_3.py`.
   When k = 1, k-NN error is 0.453667.
   When k = 3, k-NN error is 0.436333.
   When k = 5, k-NN error is 0.425667.
4. Attached as `ex1_4.py`.
   When k = 1, k-NN error is 0.455000.
   When k = 3, k-NN error is 0.440000.
   When k = 5, k-NN error is 0.428333.
5. Attached as `ex1_5.py`.
   When sigma = 0.01, Bayes error: 0.004667, 1-NN error: 0.012333.
   When sigma = 0.1, Bayes error: 0.070333, 1-NN error: 0.103667.
   When sigma = 1, Bayes error: 0.395000, 1-NN error: 0.468333.
   When sigma = 10, Bayes error: 0.496000, 1-NN error: 0.498000
   Bayes error is always smaller than 1-NN error, when σ is large, they become close.

Exercise 2

1. The likelihood function is $\prod_{i=1}^{n} \frac{\mu_i^{y_i}}{y_i!} \exp(-\mu_i)$.
   Therefore, the log-likelihood function is $\sum_{i=1}^{n} y_i \log(\mu_i) - \log(y_i!) - \mu_i$
2. $\mu_i \in [0, +\infty)$, $(\mathbf{w}^T \mathbf{x}_i + b) \in \mathbb{R}$, therefore we take the log transform on μ_i
3. Let $\log \mu_i$ equals to $\mathbf{w}^T \mathbf{x}_i + b$, then we got the maximization problem

$$\max_{\mathbf{w},b} \sum_{i=1}^{n} y_i(\mathbf{w}^T \mathbf{x}_i + b) - \log(y_i!) - \exp(\mathbf{w}^T \mathbf{x}_i + b) \tag{1}$$

   Take the negative and dropped the constants we got the minimization problem

$$\min_{\mathbf{w},b} \sum_{i=1}^{n} \exp(\mathbf{w}^T \mathbf{x}_i + b) - y_i(\mathbf{w}^T \mathbf{x}_i + b) \tag{2}$$

4. Let's denote the (2) as function $f(\mathbf{w}, b)$. Then we can calculate the gradient

$$
\begin{aligned}
\frac{\partial f}{\partial \mathbf{w}} &\quad \sum_{i=1}^{n} \mathbf{x}_i \exp\left(\mathbf{w}^T \mathbf{x}_i + b\right) - y_i \mathbf{x}_i \\
\frac{\partial f}{\partial b} &\quad \sum_{i=1}^{n} \exp\left(\mathbf{w}^T \mathbf{x}_i + b\right) - y_i
\end{aligned}
\tag{3}
$$

Algorithm 1: Gradient descent for poisson regression.

Input: $X \in \mathbb{R}^{d \times n}, \mathbf{y} \in \mathbb{R}^n, \mathbf{w}_0 \quad \mathbf{0}_d, b_0 \quad 0, \mathsf{max\_pass} \in N, \eta > 0, \mathsf{tol} > 0$

Output: $\mathbf{w}, b$

for $t \quad 1, 2, \ldots, \mathsf{max\_pass}$ **do**

  calculate the gradient $[\mathbf{d}_{\mathbf{w_t}\ \mathbf{1}}^T, \mathbf{d}_{b_t\ 1}]^T \quad [\frac{\partial f}{\partial \mathbf{w_t}\ \mathbf{1}}^T, \frac{\partial f}{\partial b_t\ 1}]^T$

  $\mathbf{w}_t \leftarrow \mathbf{w}_{t\ 1} - \eta \mathbf{d}_{\mathbf{w_t}\ \mathbf{1}}$

  $b_t \leftarrow b_{t\ 1} - \eta \mathbf{d}_{b_t\ 1}$

  if $\|\mathbf{w}_t - \mathbf{w}_{t\ 1}\| \leqslant \mathsf{tol}$ **then**

    break

$\mathbf{w} \leftarrow \mathbf{w}_t, \ b \leftarrow b_t$
