7.1

$$x_i = \begin{pmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{iD} \end{pmatrix}, \quad \mathbb{X} = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{pmatrix}, \quad \mathbb{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix}, \quad w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_D \end{pmatrix}$$

$$E(w) = -\log L(w) = -\sum_{n=1}^{N} \{ t_n \log \hat{t}_n + (1-t_n)\log(1-\hat{t}_n) \}$$

1. $$\nabla E(w) = \sum_{n=1}^{N} (\hat{t}_n - t_n) x_n = \mathbb{X}^T(\hat{\mathbb{t}} - \mathbb{t})$$ を示す。

$$\nabla E(w) = -\sum_{n=1}^{N} \left\{ t_n \frac{\hat{t}_n(1-\hat{t}_n)x_n}{\hat{t}_n} - (1-t_n)\frac{\hat{t}_n(1-\hat{t}_n)}{(1-\hat{t}_n)} x_n \right\}$$

$$= -\sum_{n=1}^{N} \{ t_n(1-\hat{t}_n)x_n - (1-t_n)\hat{t}_n x_n \}$$

$$= -\sum_{n=1}^{N} (t_n - \hat{t}_n) x_n$$

$$= \mathbb{X}^T(\hat{\mathbb{t}} - \mathbb{t})$$ //

2. $$w' = w - \eta \mathbb{X}^T(\hat{\mathbb{t}} - \mathbb{t})$$

7.3

1. ステップサイズが大きすぎるため、最適解を超えて更新してしまうため。

2. 更新回数が増加すると、実行時間が大きくなる。
目的関数の差分がある値よりも小さくなったら打ち切るという処理を実行しているが、
ステップサイズが小さすぎると、最適解に到達するよりも先に、更新が遅くて打ち切られてしまうため。

3. 勾配法とニュートン法では扱っている計算の構造が異なるため、更新回数によって両者を比較することは難しいが、最終的な ~~最終的な~~ 目的関数値、実行時間は共にニュートン法の方が小さくなっているので、ニュートン法の方が優れていると言える。
このデータセットにおいては

4. ニュートン法は、現時点での解付近において目的関数が凸二次関数で近似されると仮定し、その凸二次関数を最小にする点に解を更新していく。そのため、この仮定がある程度正しければ、ステップサイズパラメータを必要としない。

5. ニュートン法は、ヘシアン行列の逆行列を計算する必要があるため、非常に次元数が大きい問題においては、その計算コストが大きくなりすぎる場合が考えられる。そのような場合は、最急降下法の方が適していると考えられる。一方で、ヘシアン行列の逆行列が計算可能であれば、ニュートン法はステップサイズパラメータのチューニングを必要とせず、より小さい実行時間で質の高い解を求めることができると考えられる。

7.5

$$E_{cross}(w) = \sum_{i=1}^{N} \left\{ -t_i \log \hat{t}_i - (1-t_i) \log (1-\hat{t}_i) \right\} \quad \cdots ①$$

$$E_{logistic}(w) = \sum_{i=1}^{N} \log \left(1 + e^{-y_i w^T x_i} \right) \quad \cdots ③$$

① = ② を示す.

①を変形して

$$E_{cross}(w) = -\sum_{i:t_i=0}^{N} \log (1-\hat{t}_i) - \sum_{i:t_i=1}^{N} \log \hat{t}_i$$

$$= -\sum_{i:t_i=0}^{N} \log \left(1 - \sigma(w^T x_i)\right) - \sum_{i:t_i=1}^{N} \log \sigma(w^T x_i)$$

$$= -\sum_{i:t_i=0}^{N} \log \sigma(-w^T x_i) - \sum_{i:t_i=1}^{N} \log \sigma(w^T x_i)$$

$$= -\sum_{i=1}^{N} \log \sigma(y_i w^T x_i)$$

$$= -\sum_{i=1}^{N} \log \frac{1}{1+\exp(-y_i w^T x_i)}$$

$$= \sum_{i=1}^{N} \log \left\{ 1 + \exp(-y_i w^T x_i) \right\} = [②の右辺] \quad //$$