5.2

1. $l_i(\boldsymbol{w}) = \max\{0, 1 - t_i(\boldsymbol{w}^T \boldsymbol{x}_i)\}$ の劣勾配

$$\partial l_i(\boldsymbol{w}) = \begin{cases} -t_i \boldsymbol{x}_i & \text{if} \quad 1 - t_i \boldsymbol{w}^T \boldsymbol{x}_i > 0 \\ \boldsymbol{z} & \text{if} \quad 1 - t_i \boldsymbol{w}^T \boldsymbol{x}_i = 0 \\ 0 & \text{if} \quad 1 - t_i \boldsymbol{w}^T \boldsymbol{x}_i < 0 \end{cases}$$

$\boldsymbol{z}$ の第j要素 z_j は　$z_j \in [0, -t_i x_{ij}]$

2. $$\frac{\partial}{\partial \boldsymbol{w}} \frac{1}{C} \|\boldsymbol{w}\|_2^2 = \frac{2}{C} \boldsymbol{w}$$

3. $$\boldsymbol{w}^{(t)} \leftarrow \boldsymbol{w}^{(t-1)} - \eta \, \partial \, l_i(\boldsymbol{w}^{(t-1)})$$ より

$$\boldsymbol{w}^{(t)} \leftarrow \boldsymbol{w}^{(t-1)} - \begin{cases} \eta\left(-t_i \boldsymbol{x}_i + \frac{2}{C} \boldsymbol{w}\right) & \text{if} \quad 1 - t_i \boldsymbol{w}^T \boldsymbol{x}_i > 0 \\ \eta\left(\frac{2}{C} \boldsymbol{w}\right) & \text{if} \quad 1 - t_i \boldsymbol{w}^T \boldsymbol{x}_i \leq 0 \end{cases}$$

5.3

1. 違いは特に見受けられない。

2. ヒンジ損失を用いた場合、分類のための超平面は正しく学習されていると思われる。一方で、二乗損失を用いて学習した場合、決定境界付近のいくつかのサンプルにおいて、分類を正しく行うことができていないため、適切に学習されなかったと思われる。

3. 問題2において違いが生じた理由は、外れ値であると考える。ヒンジ損失は当たっているデータについては損失が 0 となり w に影響しないため、外れ値は無視される。一方で、二乗損失では、識別超平面から離れたデータは、当たっているか当たっていないかに関らず大きな損失が課せられる。そのため、外れ値が存在した場合、その影響を受け、トータルとして損失を小さくするために超平面付近に誤分類が見られたと考えられる。

5.4

1. あやめデータを予測した時、virginica ではないが、virginica であると予測してしまう状況。

2. 再現率　　リコール、true positive rate

3. 0.8

4. 不可能である。

［理由］

AUC は [0, 1] の範囲における ROC 曲線と x 軸が囲む領域の面積を表している。ある false positive rate において、$C = 10^{-1}$ におけるモデルの true positive rate が、それ以外の C におけるモデルの true positive rate を下回っていたとしていても、$C = 10^{-1}$ におけるモデルの AUC が それ以外の C における AUC を上回ることはありえると考えられるため。