FAI Homework 3

1. Let $z = y_n \mathbf{w^T x_n}$

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{w_i}}(\max(1 - y_n \mathbf{w^T x_n}, 0)) &= \frac{\partial}{\partial z}\max(1-z,0) \cdot \frac{\partial}{\partial \mathbf{w_i}} y_n \mathbf{w^T x_n} \\
&= \begin{cases} 0 & \text{, if } y_n \mathbf{w^T x_n} > 1 \\ (-1)\cdot(y_n(\mathbf{x_n})_i) & \text{, if } y_n \mathbf{w^T x_n} < 1 \\ \text{not differentiable} & \text{, if } y_n \mathbf{w^T x_n} = 1 \end{cases} \\
\frac{\partial}{\partial \mathbf{w}}(\max(1 - y_n \mathbf{w^T x_n}, 0)) &= \sum_{i=0}^{d} \frac{\partial}{\partial \mathbf{w_i}}(\max(1 - y_n \mathbf{w^T x_n}, 0)) \\
&= \begin{cases} \mathbf{0} & \text{, if } y_n \mathbf{w^T x_n} > 1 \\ -y_n \mathbf{x_n} & \text{, if } y_n \mathbf{w^T x_n} < 1 \\ \text{not differentiable} & \text{, if } y_n \mathbf{w^T x_n} = 1 \end{cases} \\
\frac{\partial}{\partial \mathbf{w}}(\max(1 - y_n \mathbf{w^T x_n}, 0))^2 &= 2(\max(1 - y_n \mathbf{w^T x_n}, 0)) \cdot \frac{\partial}{\partial \mathbf{w}}(\max(1 - y_n \mathbf{w^T x_n}, 0)) \\
&= \begin{cases} \mathbf{0} & \text{, if } y_n \mathbf{w^T x_n} \geq 1 \\ 2(1 - y_n \mathbf{w^T x_n})(-y_n \mathbf{x_n}) & \text{, if } y_n \mathbf{w^T x_n} < 1 \end{cases} \\
&= 2\max(1 - y_n \mathbf{w^T x_n}, 0)(-y_n \mathbf{x_n}) \\
\nabla E_{in}(\mathbf{w}) &= \frac{1}{N}\sum_{n=1}^{N} \frac{\partial}{\partial \mathbf{w}}(\max(1 - y_n \mathbf{w^T x_n}, 0))^2 \\
&= \frac{2}{N}\sum_{n=1}^{N} \max(1 - y_n \mathbf{w^T x_n}, 0)(-y_n \mathbf{x_n})
\end{aligned}
$$

2.

$$
\begin{aligned}
\mathbf{w}_{\text{lin}} &= (X^T X)^{-1} X^T \mathbf{y} \\
&= (V(\Sigma^T U^T U \Sigma)V^T)^{-1}(V\Sigma^T U^T)\mathbf{y} \\
&= V(\Sigma^T U^T U \Sigma)^{-1} V^T V \Sigma^T U^T \mathbf{y} \\
&= V(\Sigma^T \Sigma)^{-1} \Sigma^T U^T \mathbf{y} \\
&= V\Gamma\Gamma^T \Sigma^T U^T \mathbf{y} \\
&= V\Gamma U^T \mathbf{y}
\end{aligned}
$$

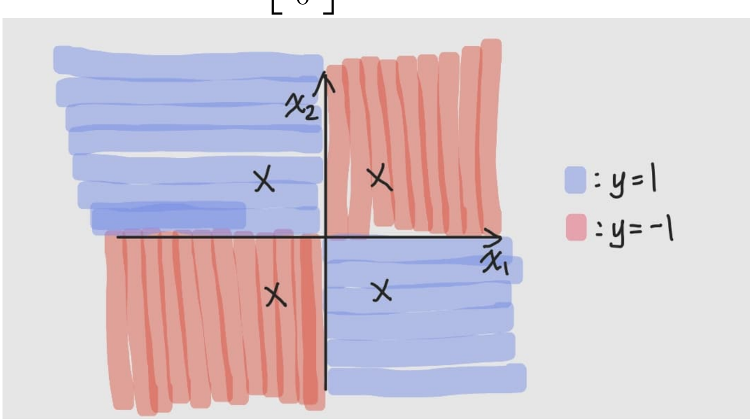
3. https://xavierbourretsicotte.github.io/MLE_Multivariate_Gaussian.html
https://en.wikipedia.org/wiki/Multivariate_normal_distribution

$$
\begin{aligned}
Let\ l(\mathbf{u}) &= \log(\prod_{n=1}^{N} p_{\mathbf{u}}(\mathbf{x}_n)) \\
&= \sum_{n=1}^{N} \log(\frac{1}{(2\pi)^{\frac{d}{2}}|I|^{\frac{1}{2}}} \exp(-\frac{1}{2}(\mathbf{x}_n - \mathbf{u})^{\mathbf{T}} I^{-1}(\mathbf{x}_n - \mathbf{u}))) \\
&= -\frac{Nd}{2}\log(2\pi) - \frac{N}{2}\log(|I|) - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \mathbf{u})^{\mathbf{T}} I^{-1}(\mathbf{x}_n - \mathbf{u})
\end{aligned}
$$

With $\frac{\partial \mathbf{w}^{\mathbf{T}}\mathbf{A}\mathbf{w}}{\partial \mathbf{w}} = 2\mathbf{A}\mathbf{w}$ if $\mathbf{w}$ is not dependent on $\mathbf{A}$ and $\mathbf{A}$ is symmetric, we have:

$$
\begin{aligned}
\frac{\partial l(\mathbf{u})}{\partial \mathbf{u}} &= \sum_{n=1}^{N} I^{-1}(\mathbf{u} - \mathbf{x}_n) = 0 \\
\Rightarrow N\mathbf{u} - \sum_{n=1}^{N}\mathbf{x}_n &= 0 \\
\Rightarrow u^* &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n
\end{aligned}
$$

4. A perception $\tilde{\mathbf{w}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}$ can separate the data.



1.

Model	accuracy	F1
Logistic Regression	0.869	0.736
Decision Tree	0.781	0.594
Random Forest	0.838	0.637

From the figure above, Logistic Regression outperforms other models.

2.

num_tree	accuracy	F1
5	0.825	0.611
11	0.838	0.600
17	0.844	0.629

From the figure above, a Random Forest with 17 trees outperforms the others.