1,

a. LFD Problem 4.8

$$E_{aug}(w) = E_{in}(w) + \lambda w^T w \quad (1)$$

$$w^T w = \sum_{i=0}^{n} w_i^2 \rightarrow \frac{\partial(w^T w)}{\partial w} = 2w \quad (2)$$

$$(1)(2) \rightarrow \nabla E_{aug}(w) = \nabla E_{in}(w) + 2\lambda w(t)$$

$$\rightarrow w(t+1) \leftarrow w(t) - \eta \nabla E_{aug}(w(t))$$

$$\leftarrow w(t) - \eta(\nabla E_{in}(w(t) + 2\lambda w(t))$$

$$\leftarrow (1-2\eta\lambda) w(t) - \eta \nabla E_{in}(w(t))$$

b, L_1 : $E_{aug}(w) = E_{in}(w) + \lambda \|w\|_1$

$$\rightarrow \nabla E_{aug}(w) = \nabla E_{in}(w) + \lambda\left(\frac{\partial \|w\|_1}{\partial w}\right)$$

$$= \nabla E_{in}(w) + \lambda \operatorname{sign}(w)$$

$$\hookrightarrow w(t+1) \leftarrow w(t) - \eta \nabla E_{aug}(w(t))$$

$$\leftarrow w(t) - \eta(\nabla E_{in}(w) + \lambda \operatorname{sign}(w))$$

c,

L1 Regularization:

Regularization Strength	Classification Error on test set	Number of 0s in learned weight vector
0	0.1028	8
0.0001	0.09813	8
0.001	0.09346	14
0.005	0.08879	17
0.01	0.07944	20
0.05	0.1028	33
0.1	0.135514	39

L2 Regularization:

Regularization Strength	Classification Error on test set	Number of 0s in learned weight vector
0	0.1028	8
0.0001	0.1028	8
0.001	0.09346	8
0.005	0.09813	8
0.01	0.09813	8
0.05	0.1168	8
0.1	0.1215	8

- Number of 0's in learned weight vector for L1 regularization increases & stays constant (8) for L2 regularization as we increase the regularization strength.

- Classification Error on Test Set for both L1 and L2 regularizations varies (decrease & increases back up around $\lambda \approx 0.05$)

2, LFD Exercise 4.5

a, $\Gamma = I$, where I is the identity matrix

$$\hookrightarrow w^T \Gamma^T \Gamma w = w^T w = \sum_{q=0}^{Q} w_q^2 \leqslant C$$

b, $\Gamma = [1 \ 1 \ \ldots \ 1] \rightarrow w^T \Gamma^T = \sum_{q=0}^{Q} w_q$

$$\hookrightarrow w^T \Gamma^T \Gamma w = \left(\sum_{q=0}^{Q} w_q \right)^2 \leqslant C$$

3, LFD Problem 4.25 (a) to (c)

a, No because the learner with minimum $E_{val}(m)$ does not guarantee minimum E_{out} since $O\left(\sqrt{\frac{\ln M}{2K_m}}\right)$ varies in each m learner

b, Yes because if all models are validated on the same validation set then K_m is the same

↳ Selecting minimum Eval will lead to minimum Eout

c, $$Me^{-2\epsilon^2 k(\epsilon)} = Me^{\ln\left(\frac{1}{M}\sum_{m=1}^{M} e^{-2\epsilon^2 K_m}\right)} = \sum_{m=1}^{M} e^{-2\epsilon^2 K_m} \quad ①$$

$$P[E_{out}(m^*) - E_{val}(m^*) > \epsilon] \leq$$
$$P[E_{out}(1) - E_{val}(1) > \epsilon] + \dots + P[E_{out}(M) - E_{val}(M > \epsilon]$$
$$\leq \sum_{m=1}^{M} e^{-2\epsilon^2 K_m} \quad ②$$

①② → $P[E_{out}(m^*) - E_{val}(m^*) > \epsilon] \leq Me^{-2\epsilon^2 k(\epsilon)}$

4, LFD Problem 5.4

a,

(i) We went wrong when we used $N = 12500$ while M is fixed at 500 for today's stock only. Moreover, those 500 stocks might have been used in the total 50,000 stocks (data snooping)

(ii) M should be 50000

$$\hookrightarrow P[|E_{in} - E_{out}| > 0.02] \leq 2 \times 50000 \times e^{-2 \times 12500 \times 0.02^2} \simeq 4.54$$

b,

(i) We only took into consideration the 500 stocks selected beforehand. Moreover, we only looked at the companies that didn't go bankrupt & stop trading

(ii) We can say something about the performance of buy & hold trading if we consider all 50,000 stocks and only companies that didn't go bankrupt or stop trading