1.

1.

a. $z^t = Wx^t$, where $x^t \in R^D$, $z^t \in R^d$, $W \in R^{d \times D}$

and $WW^T = I_d$

since $v^t = W^T z^t = W^T W x^t$

$W^T W = WW^T = I_d \neq I_D$

Therefore, $v^t = W^T W x^t \neq x^t$

Professor HighLowHigh's claim is (wrong).

b. $\sum_{t=1}^{N} \|x^t - v^t\|_2^2 = \sum_{t=1}^{N} ((x^t - v^t)^T (x^t - v^t))$

$= \sum_{t=1}^{N} ((x^t)^T x^t - (x^t)^T v^t - (v^t)^T x^t + (v^t)^T v^t)$

$= \sum_{t=1}^{N} ((x^t)^T x^t + (v^t)^T v^t - (x^t)^T W^T z^t - (W^T z^t)^T x^t)$

$= \sum_{t=1}^{N} ((x^t)^T x^t + (v^t)^T v^t - (x^t W)^T z^t - (z^t)^T W x^t)$

$= \sum_{t=1}^{N} ((x^t)^T x^t + (v^t)^T v^t - (z^t)^T z^t - (z^t)^T z^t)$

$= \sum_{t=1}^{N} ((x^t)^T x^t + (v^t)^T v^t - 2(z^t)^T z^t)$

$= \sum_{t=1}^{N} ((x^t)^T x^t + (v^t)^T v^t - 2(z^t)^T W W^T z^t)$

$= \sum_{t=1}^{N} ((x^t)^T x^t + (v^t)^T v^t - 2(W^T z^t)^T W^T z^t)$

$= \sum_{t=1}^{N} ((x^t)^T x^t + (v^t)^T v^t - 2(v^t)^T v^t)$

$= \sum_{t=1}^{N} ((x^t)^T x^t - (v^t)^T v^t)$

$= \sum_{t=1}^{N} \|x^t\|_2^2 - \sum_{t=1}^{N} \|v^t\|_2^2$

Therefore, Professor HighLowHigh's claim is right.

2.

$z_h^t = g(a_h^t) = g\left(\sum_{j=1}^{d} w_{h,j} x_j^t + w_{h,0}\right)$, $y_i^t = g(a_i^t) = g\left(\sum_{h=1}^{H+1} v_{i,h} z_h^t + v_{i,0}\right)$

a.

$$\Delta v_{i,h} = -\eta \frac{\partial E(W,V|z)}{\partial v_{i,h}}$$

$$= -\eta \frac{\partial E(W,V|z)}{\partial y_i^t} \cdot \frac{\partial g(a_i^t)}{\partial a_i^t} \cdot \frac{\partial a_i^t}{\partial v_{i,h}}$$

$$= -\eta \frac{\partial L(r_i^t, y_i^t)}{\partial y_i^t} \cdot \frac{\partial g(a_i^t)}{\partial a_i^t} \cdot \frac{\partial \sum_{h=1}^{H+1} v_{i,h} z_h^t + w_{i,0}}{\partial v_{i,h}}$$

$$= -\eta \frac{\partial L(r_i^t, y_i^t)}{\partial y_i^t} g'(a_i^t) \cdot z_h^t$$

$$= \eta \delta_i^t z_h^t, \text{ where } \delta_i^t = g'(a_i^t)\left(-\frac{\partial L(r_i^t, y_i^t)}{\partial y_i^t}\right)$$

b.

$$\Delta w_{h,j} = -\eta \delta \frac{\partial E(W,V|z)}{\partial h,j}$$

$$= -\eta \frac{\partial E(W,V|z)}{\partial y_i^t} \times \frac{\partial y_i^t}{\partial a_i^t} \times \frac{\partial a_i^t}{\partial z_h^t} \times \frac{\partial z_h^t}{\partial a_h^t} \times \frac{\partial a_h^t}{\partial w_{h,j}^t}$$

$$= -\eta \sum_{i=1}^{K} \left(\frac{\partial L(h_i^t, y_i^t)}{\partial y_i^t} \times \frac{\partial g(a_i^t)}{\partial a_i^t} \times \frac{\partial \sum_{h=1}^{H} (v_{i,h} z_h^t + v_{i,0})}{\partial z_h^t}\right)$$

$$\times \frac{\partial g(a_h^t)}{\partial a_h^t} \cdot \frac{\partial \sum_{j=1}^{d} w_{h,j} x_j^t + w_0)}{\partial w_{h,j}}$$

$$= -\eta \left(\sum_{i=1}^{K} \frac{\partial L(h_i^t, y_i^t)}{\partial y_i^t} \times g'(a_i^t) \times v_{i,h}^t\right) \times g'(a_h^t) \times x_j^t$$

$$\delta_i^t = g'(a_i^t)\left(-\frac{\partial L(h_i^t, y_i^t)}{\partial y_i^t}\right)$$

$$\delta_h^t = g'(a_h^t)\left(\sum_{i=1}^{K} \delta_i^t v_{i,h}\right)$$

$$\therefore \Delta w_{h,j} = \eta \delta_h^t x_j^t, \text{ where } \delta_h^t = g'(a_h^t)\left(\sum_{i=1}^{t} \delta_i^t v_{i,h}\right)$$

3.

Error rates for MySVM2 with m=40 for Boston50						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.257425742574257 43	0.1485148 5148514854	0.1386138 6138613863	0.3267326 732673267	0.4356435 643564357	0.2613861 386138614	0.1117716 1023105017

Error rates for MySVM2 with m=200 for Boston50						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.435643564356435 7	0.3069306 93069307	0.4455445 544554455	0.3168316 831683168	0.1782178 217821782 7	0.3366336 633663366 6	0.0980148 013525907 3

Error rates for MySVM2 with m=n for Boston50						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.455445544554455 4	0.4950495 049504950 5	0.1089108 910891089	0.5940594 05940594	0.8217821 782178218	0.4950495 049504951 6	0.2311833 484783567 5

Error rates for LogisticRegression with Boston50						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.137254901960784 27	0.1485148 5148514854	0.2277227 722772277	0.0099009 900990099 1	0.2574257 425742574 3	0.1561638 516792855 6	0.0862363 379624220 4

Error rates for MySVM2 with m=40 for Boston25						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.1683168	0.1287128	0.1386138	0.0792079	0.2277227	0.1485148	0.0489072

3168316836	7128712872	6138613863	2079207917	722772277	514851485	8330783553

Error rates for MySVM2 with m=200 for Boston25						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.168316831683168336	0.158415841584155	0.0	0.089108910891088	0.386138613861385	0.160396039604	0.127933657838661

Error rates for MySVM2 with m=n for Boston25						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.168316831683168336	0.178217821782177	0.0	0.336633663366336	0.564356435643564	0.249504950495056	0.190078381719072623

Error rates for LogisticRegression with Boston25						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.225490196078431335	0.254901960784313737	0.089108910891088	0.049504950495055	0.339999999999997	0.191801203649776744	0.107585749665441222

I found that the result of Boston25 is better than Boston50.