CS2109S Tutorial 10

AY 25/26 Sem 1 — github/omgeta

A.

1. $w^\top x = -0.1 < 0$ so SVM classifies x as Not-VIP

2. We need 17 support vectors, each of 30 attributes, and 17 α coefficients, giving total cost of $17 \times 30 + 17 = 527$

B.

1. Total variance $= \frac{6^2+4.5^2+2^2+1^2+0.5^2}{n-1} = \frac{61.5}{n-1}$
   Keeping $6, 4$ gives variance $\frac{6^2+4^2}{61.5} = 0.915 > 0.9$ so suffice to keep just them.

2. For compressing $64 \times 64 = 4096$ mean-centered features to 50 features, we need compression matrix $\tilde{U} \in \mathbb{R}^{4096\times 50}$, the 4096 feature means for normalisation and denormalisation. For decompression of the original image, we already have $\tilde{U}$, so we only need the additional compressed images of $50 \times 500 = 25000$ values. So in total we have $(4096 \times 50) + 25000 + 4096 = 233896$.

C.

1. $\hat{y} = \text{ReLU}(W^{[2]\top}(W^{[1]\top}\begin{bmatrix}1\\1\end{bmatrix})) = \text{ReLU}(W^{[2]\top}\begin{bmatrix}3\\2\end{bmatrix}) = \text{ReLU}(8) = 8$

2. Note $J(W) = L_1 + L_2 = \frac{1}{2\times 2}((\hat{y}^{(1)} - y^{(1)})^2 + (\hat{y}^{(2)} - y^{(2)})^2)$ so:

$$\frac{\partial J(W)}{\partial W_{11}^{[1]}} = \frac{\partial L_1}{\partial W_{11}^{[1]}} + \frac{\partial L_2}{\partial W_{11}^{[1]}}$$

   Note also for any sample i, $\hat{y}^{(i)}$ predicted as:

$$L_i = \frac{1}{4}(\hat{y}^{(i)} - y^{(i)})^2 \implies \frac{\partial L_i}{\partial \hat{y}^{(i)}} = \frac{1}{2}(\hat{y}^{(i)} - y^{(i)})$$

$$\hat{y}^{(i)} = \text{ReLU}(f^{(i)}) \implies \frac{\partial y^{(i)}}{\partial f^{(i)}} = 1_{f^{i)}>0}$$

$$f^{(i)} = W_{11}^{[2]}a_1^{(i)} + W_{21}^{a_2^{(i)}} \implies \frac{\partial f^{(i)}}{\partial a_1^{(i)}} = W_{11}^{[2]}$$

$$a_1^{(i)} = W_{11}^{[1]}x_1^{(i)} + W_{21}^{[1]}x_2 \implies \frac{\partial a_1^{(i)}}{\partial W_{11}^{[1]}} = x_1^{(i)}$$

   Then by chain rule, $\frac{\partial L_i}{\partial W_{11}^{[1]}} = \frac{1}{2}(\hat{y}^{(i)} - y^{(i)})1_{f^{i)}>0}W_{11}^{[2]}x_1^{(i)}$ so $\frac{\partial J(W)}{\partial W_{11}^{[1]}} = 4 + 0 = 4$

D.

1. Kernel size $= 2 \times 2 \times 3$ $(H \times W \times C)$, with 5 kernels to produce 5-channel output.
   Therefore, total parameters $= 2 \times 2 \times 3 \times 5 = 60$

2. Kernel size $= 1 \times 1 \times 3$ with all 3 weights as $\frac{1}{3}$, stride $= 1$, padding $= 0$