Module 13

Soft Margin SVM
Solutions

Problem 1: 25 points

Given an $n \times n$ matrix A, recall that its adjugate $\widetilde{A} = (b_{ij})$ is given by

$$b_{ij} = (-1)^{i+j} \det(A_{ji}),$$

where A_{ji} is the $(n-1) \times (n-1)$ matrix obtained from A by deleting its jth row and its i-column. For example, when $n = 2$, if

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

then

$$\widetilde{A} = \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

(1) When $n = 2$, check directly that

$$\widetilde{A}A = \det(A)I_2.$$

(2) For any matrix

$$H = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix},$$

prove that

$$\det(A + H) - \det(A) = \operatorname{tr}(\widetilde{A}H) + \det(H).$$

(3) Recall that the Frobenius norm of H is given by

$$\|H\|_F = \sqrt{h_{11}^2 + h_{12}^2 + h_{21}^2 + h_{22}^2}.$$

Check that for $H \neq 0$,

$$\frac{|h_{ij}|}{\|H\|_F} \leq 1,$$

and deduce from this that

$$\lim_{\|H\|_F \mapsto 0} \frac{\det(H)}{\|H\|_F} = 0.$$

If we write

$$\epsilon(H) = \frac{\det(H)}{\|H\|_F},$$

then deduce that

$$\det(A+H) - \det(A) = \mathrm{tr}(\widetilde{A}H) + \|H\|_F\,\epsilon(H)$$

with $\lim_{\|H\|_F \mapsto 0} \epsilon(H) = 0$ $(H \neq 0)$.

Conclude that the derivative $d\det_A(H)$ of the determinant function $\det\colon \mathrm{M}_2(\mathbb{R}) \to \mathbb{R}$ at A applied to H is given by

$$d\det_A(H) = \mathrm{tr}(\widetilde{A}H).$$

(4) If A is invertible, prove that

$$d\det_A(H) = \det(A)\mathrm{tr}(A^{-1}H).$$

Deduce that the derivative of the function $f\colon \mathrm{M}_2(\mathbb{R}) \to \mathbb{R}$ given by $f(A) = \log\det(A)$ is given by

$$df_A(H) = \mathrm{tr}(A^{-1}H).$$

Solution. (1) We have

$$\widetilde{A}A = \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = (a_{11}a_{22} - a_{12}a_{21}) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(2) We have

$$\begin{aligned}
\det(A+H) &= \begin{vmatrix} a_{11}+h_{11} & a_{12}+h_{12} \\ a_{21}+h_{21} & a_{22}+h_{22} \end{vmatrix} \\
&= (a_{11}+h_{11})(a_{22}+h_{22}) - (a_{12}+h_{12})(a_{21}+h_{21}) \\
&= a_{11}h_{22} + a_{22}h_{11} + a_{11}a_{22} + h_{11}h_{22} \\
&= \quad -(a_{12}h_{21} + a_{21}h_{12} + a_{12}a_{21} + h_{12}h_{21}) \\
&= a_{11}h_{22} + a_{22}h_{11} - a_{12}h_{21} - a_{21}h_{12} \\
&\quad + a_{11}a_{22} - a_{12}a_{21} + h_{11}h_{22} - h_{12}h_{21},
\end{aligned}$$

and

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21},$$

so we get

$$\det(A+H) - \det(A) = a_{11}h_{22} + a_{22}h_{11} - a_{12}h_{21} - a_{21}h_{12} + h_{11}h_{22} - h_{12}h_{21}.$$

Since

$$\operatorname{tr}(\widetilde{A}H) = \operatorname{tr}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}\begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} = \operatorname{tr}\begin{pmatrix} a_{22}h_{11} - a_{12}h_{21} & a_{22}h_{12} - a_{12}h_{22} \\ a_{11}h_{21} - a_{21}h_{11} & a_{11}h_{22} - a_{21}h_{12} \end{pmatrix}$$
$$= a_{11}h_{22} + a_{22}h_{11} - a_{12}h_{21} - a_{21}h_{12}$$

and

$$\det(H) = h_{11}h_{22} - h_{12}h_{21},$$

we obtain

$$\det(A + H) - \det(A) = \operatorname{tr}(\widetilde{A}H) + \det(H).$$

(3) We have

$$\frac{|h_{ij}|}{\|H\|_F} = \frac{|h_{ij}|}{\sqrt{h_{11}^2 + h_{12}^2 + h_{21}^2 + h_{22}^2}} \leq 1,$$

since

$$h_{ij}^2 \leq h_{11}^2 + h_{12}^2 + h_{21}^2 + h_{22}^2.$$

If $H \neq 0$, since we can write

$$\begin{aligned}\frac{\det(H)}{\|H\|_F} &= \frac{h_{11}h_{22} - h_{12}h_{21}}{\sqrt{h_{11}^2 + h_{12}^2 + h_{21}^2 + h_{22}^2}} \\ &= \frac{h_{11}}{\sqrt{h_{11}^2 + h_{12}^2 + h_{21}^2 + h_{22}^2}}\, h_{22} - \frac{h_{12}}{\sqrt{h_{11}^2 + h_{12}^2 + h_{21}^2 + h_{22}^2}}\, h_{21}.\end{aligned}$$

Since

$$\frac{|h_{11}|}{\sqrt{h_{11}^2 + h_{12}^2 + h_{21}^2 + h_{22}^2}} \leq 1 \quad \text{and} \quad \frac{|h_{12}|}{\sqrt{h_{11}^2 + h_{12}^2 + h_{21}^2 + h_{22}^2}} \leq 1,$$

we deduce that

$$\frac{|\det(H)|}{\|H\|_F} \leq |h_{22}| + |h_{21}|,$$

so if $\|H\|_F = \sqrt{h_{11}^2 + h_{12}^2 + h_{21}^2 + h_{22}^2}$ tends to 0, then all the h_{ij} then to zero and

$$\lim_{\|H\|_F \mapsto 0} \frac{|\det(H)|}{\|H\|_F} = 0,$$

which implies that

$$\lim_{\|H\|_F \mapsto 0} \frac{\det(H)}{\|H\|_F} = 0.$$

If we write

$$\epsilon(H) = \frac{\det(H)}{\|H\|_F},$$

then we deduce from

$$\det(A+H) - \det(A) = \mathrm{tr}(\widetilde{A}H) + \det(H)$$

that

$$\det(A+H) - \det(A) = \mathrm{tr}(\widetilde{A}H) + \|H\|_F\, \epsilon(H)$$

with $\lim_{\|H\|_F \mapsto 0} \epsilon(H) = 0$ $(H \neq 0)$.

The above is exactly the definition of the derivative $d\det_A(H)$ of the determinant function $\det\colon \mathrm{M}_2(\mathbb{R}) \to \mathbb{R}$ at A applied to H, with

$$d\det_A(H) = \mathrm{tr}(\widetilde{A}H).$$

(4) If A is invertible, then $\det(A) \neq 0$ and from (1) we know that

$$\widetilde{A} = \det(A)A^{-1},$$

so by the linearity of the trace function

$$\mathrm{tr}(\widetilde{A}H) = \mathrm{tr}(\det(A)A^{-1}H) = \det(A)\mathrm{tr}(A^{-1}H),$$

and we have

$$d\det_A(H) = \det(A)\mathrm{tr}(A^{-1}H).$$

By using the chain rule

$$df_A = d(\log \circ \det)_A = d\log_{\det(A)} \circ d\det_A,$$

and since

$$d\log_z = \frac{1}{z},$$

we get

$$df_A(H) = \frac{1}{\det(A)} d\det_A(H) = \frac{1}{\det(A)} \det(A)\mathrm{tr}(A^{-1}H) = \mathrm{tr}(A^{-1}H).$$