Appendix B Vector and matrix formulas

This appendix lists some useful vector and matrix formulas. Note that these formulas are selected for the purpose of deriving equations in this book in Bayesian speech and language processing, and do not cover the whole field of vector and matrix formulas.

B.1 Trace

$$\mathrm{tr}[a] = a, \tag{B.1}$$

$$\mathrm{tr}[\mathbf{ABC}] = \mathrm{tr}[\mathbf{BCA}] = \mathrm{tr}[\mathbf{CAB}], \tag{B.2}$$

$$\mathrm{tr}[\mathbf{A} + \mathbf{B}] = \mathrm{tr}[\mathbf{A}] + \mathrm{tr}[\mathbf{B}], \tag{B.3}$$

$$\mathrm{tr}[\mathbf{A}^\intercal] = \mathrm{tr}[\mathbf{A}], \tag{B.4}$$

$$\mathrm{tr}[\mathbf{A}(\mathbf{B} + \mathbf{C})] = \mathrm{tr}[\mathbf{AB} + \mathbf{AC}]. \tag{B.5}$$

B.2 Transpose

$$(\mathbf{ABC})^\intercal = \mathbf{C}^\intercal \mathbf{B}^\intercal \mathbf{A}^\intercal, \tag{B.6}$$

$$(\mathbf{A} + \mathbf{B})^\intercal = \mathbf{A}^\intercal + \mathbf{B}^\intercal. \tag{B.7}$$

B.3 Derivative

$$\frac{\partial \log |\mathbf{A}|}{\partial \mathbf{A}} = (\mathbf{A}^\intercal)^{-1}, \tag{B.8}$$

$$\frac{\partial \mathbf{a}^\intercal \mathbf{b}}{\partial \mathbf{a}} = \frac{\partial \mathbf{b}^\intercal \mathbf{a}}{\partial \mathbf{a}} = \mathbf{b}, \tag{B.9}$$

$$\frac{\partial \mathbf{a}^\intercal \mathbf{C} \mathbf{b}}{\partial \mathbf{C}} = \mathbf{a}\mathbf{b}^\intercal, \tag{B.10}$$

$$\frac{\partial \mathrm{tr}(\mathbf{AB})}{\partial \mathbf{A}} = \mathbf{B}^\intercal, \tag{B.11}$$

$$\frac{\partial \mathrm{tr}(\mathbf{ACB})}{\partial \mathbf{C}} = \mathbf{A}^\mathsf{T}\mathbf{B}^\mathsf{T}, \tag{B.12}$$

$$\frac{\partial \mathrm{tr}(\mathbf{AC}^{-1}\mathbf{B})}{\partial \mathbf{C}} = -(\mathbf{C}^{-1}\mathbf{BAC}^{-1})^\mathsf{T}, \tag{B.13}$$

$$\frac{\partial \mathbf{a}^\mathsf{T}\mathbf{Ca}}{\partial \mathbf{a}} = (\mathbf{C} + \mathbf{C}^\mathsf{T})\mathbf{a}, \tag{B.14}$$

$$\frac{\partial \mathbf{b}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{DAc}}{\partial \mathbf{A}} = \mathbf{D}^\mathsf{T}\mathbf{Abc}^\mathsf{T} + \mathbf{DAcb}^\mathsf{T}. \tag{B.15}$$

B.4 Complete square

When $\mathbf{A}$ is a symmetric matrix,

$$\mathbf{x}^\mathsf{T}\mathbf{Ax} - 2\mathbf{x}^\mathsf{T}\mathbf{b} + c = (\mathbf{x} - \mathbf{u})^\mathsf{T}\,\mathbf{A}\,(\mathbf{x} - \mathbf{u}) + v, \tag{B.16}$$

where

$$\begin{aligned} \mathbf{u} &\triangleq \mathbf{A}^{-1}\mathbf{b} \\ v &\triangleq c - \mathbf{b}^\mathsf{T}\mathbf{A}^{-1}\mathbf{b}. \end{aligned} \tag{B.17}$$

By using the above complete square formula, we can also derive the following formula when matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are symmetric:

$$\begin{aligned} &(\mathbf{x} - \mathbf{b}_1)^\mathsf{T}\mathbf{A}_1(\mathbf{x} - \mathbf{b}_1) + (\mathbf{x} - \mathbf{b}_2)^\mathsf{T}\mathbf{A}_2(\mathbf{x} - \mathbf{b}_2) \\ &\quad = \mathbf{x}^\mathsf{T}\underbrace{(\mathbf{A}_1 + \mathbf{A}_2)}_{\triangleq \mathbf{A}}\mathbf{x} - 2\mathbf{x}^\mathsf{T}\underbrace{(\mathbf{A}_1\mathbf{b}_1 + \mathbf{A}_2\mathbf{b}_2)}_{\triangleq \mathbf{b}} + \underbrace{\mathbf{b}_1^\mathsf{T}\mathbf{A}_1\mathbf{b}_1 + \mathbf{b}_2^\mathsf{T}\mathbf{A}_2\mathbf{b}_2}_{\triangleq c} \\ &\quad = (\mathbf{x} - \mathbf{u})^\mathsf{T}\,(\mathbf{A}_1 + \mathbf{A}_2)\,(\mathbf{x} - \mathbf{u}) + v, \end{aligned} \tag{B.18}$$

where

$$\begin{aligned} \mathbf{u} &= (\mathbf{A}_1 + \mathbf{A}_2)^{-1}(\mathbf{A}_1\mathbf{b}_1 + \mathbf{A}_2\mathbf{b}_2), \\ v &= \mathbf{b}_1^\mathsf{T}\mathbf{A}_1\mathbf{b}_1 + \mathbf{b}_2^\mathsf{T}\mathbf{A}_2\mathbf{b}_2 - (\mathbf{A}_1\mathbf{b}_1 + \mathbf{A}_2\mathbf{b}_2)^\mathsf{T}(\mathbf{A}_1 + \mathbf{A}_2)^{-1}(\mathbf{A}_1\mathbf{b}_1 + \mathbf{A}_2\mathbf{b}_2). \end{aligned} \tag{B.19}$$

B.5 Woodbury matrix inversion

$$(\mathbf{A} + \mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}. \tag{B.20}$$

https://doi.org/10.1017/CBO9781107295360.011 Published online by Cambridge University Press