HUT , Institute of mathematics — Gripenberg
Mat-1.196 Mathematics of neural networks
Exercise 10
19.3–5.4.2002

1. Show that if C is a $d\times d$ symmetric matrix with nonnnegative eigenvalues, and V_0 is a $m\times d$ matrix, then the matrix $I-V_0V_0^{\mathrm{T}}+V_0\mathrm{e}^{Ct}V_0^{\mathrm{T}}$ is invertible for all $t\ge 0$.

Solution: Since C is symmetric there is an orthogonal matrix U such that $C=U\Lambda U^{\mathrm{T}}$ where Λ is a diagonal matrix with the eigenvalues of C on the diagonal. Furthermore, it then follows that $\mathrm{e}^{Ct}=U\mathrm{e}^{\Lambda t}U^{\mathrm{T}}$. Since U is orthogonal we have $UU^{\mathrm{T}}=I$ so that

$$I-V_0V_0^{\mathrm{T}}+V_0\mathrm{e}^{Ct}V_0^{\mathrm{T}}=I+V_0U\big(\mathrm{e}^{\Lambda t}-I\big)U^{\mathrm{T}}V_0^{\mathrm{T}}.$$

Since $\mathrm{e}^{\lambda t}-1\ge 0$ for all $t\ge 0$ when $\lambda\ge 0$, we conclude that if X is a column vector in $\mathbb{R}^m$ and $Y=U^{\mathrm{T}}V_0^{\mathrm{T}}x$ then

$$Y^{\mathrm{T}}\big(\mathrm{e}^{\Lambda t}-I\big)Y\ge 0,$$

so that

$$X\Big(I-V_0V_0^{\mathrm{T}}+V_0\mathrm{e}^{Ct}V_0^{\mathrm{T}}\Big)X\ge X^{\mathrm{T}}X,\quad t\ge 0.$$

This inequality gives the desired conclusion, because if a matrix A is not invertible, then there is a nonzero vector X such that $AX=0$ and then $X^{\mathrm{T}}AX=0$.

---

2. Suppose that C is a $d\times d$ symmetric matrix with nonnegative eigenvalues, and that V_0 is a $m\times d$ matrix. Show that

$$P(t)=\mathrm{e}^{Ct}V_0^{\mathrm{T}}(I-V_0V_0^{\mathrm{T}}+V_0\mathrm{e}^{2Ct}V_0^{\mathrm{T}})^{-1}V_0\mathrm{e}^{Ct},\quad t\ge 0,$$

is the solution to the equation

$$P'(t)=P(t)C+CP(t)-2P(t)CP(t),\quad P(0)=V_0^{\mathrm{T}}V_0,\quad t\ge 0.$$

Solution: First we observe that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{Ct}=C\mathrm{e}^{Ct}=\mathrm{e}^{Ct}C,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{2Ct}=2\mathrm{e}^{Ct}C\mathrm{e}^{Ct},$$

and that

$$\frac{\mathrm{d}}{\mathrm{d}t}A(t)^{-1}=-A(t)^{-1}A'(t)A^{-1}(t).$$

Using these results and the definition of $P(t)$ we conclude that

$$\begin{aligned}P'(t)&=C\mathrm{e}^{Ct}V_0^{\mathrm{T}}(I-V_0V_0^{\mathrm{T}}+V_0\mathrm{e}^{2Ct}V_0^{\mathrm{T}})^{-1}V_0\mathrm{e}^{Ct}\\&\quad-\mathrm{e}^{Ct}V_0^{\mathrm{T}}(I-V_0V_0^{\mathrm{T}}+V_0\mathrm{e}^{2Ct}V_0^{\mathrm{T}})^{-1}V_0\mathrm{e}^{Ct}C\mathrm{e}^{Ct}V_0^{\mathrm{T}}(I-V_0V_0^{\mathrm{T}}+V_0\mathrm{e}^{2Ct}V_0^{\mathrm{T}})^{-1}V_0\mathrm{e}^{Ct}\\&\quad+\mathrm{e}^{Ct}V_0^{\mathrm{T}}(I-V_0V_0^{\mathrm{T}}+V_0\mathrm{e}^{2Ct}V_0^{\mathrm{T}})^{-1}V_0\mathrm{e}^{Ct}C=CP(t)-2P(t)CP(t)+P(t)C,\end{aligned}$$

which is what we wanted to prove.

---

3. Construct an neural network and an algorithm for updating the weights and thresholds such that if the inputs $\mathbf{x}_n$ are (independent) random vectors with expectation value $E(\mathbf{x})$, then the output $\mathbf{y}_n \approx \mathbf{x}_n - E(\mathbf{x})$ when $n \to \infty$.

Solution: We take $L = 1$, (no hidden layer), $\sigma_1(\underline{t}) = \underline{t}$ and $W_1 = I$, the identity matrix. Thus the problem is to calculate the thresholds so that they converge to the expectation value. One possibility is to take $\tau_0 = 0$

$$\tau_{n+1} = \tau_n + \frac{1}{n}(\mathbf{x}_n - \tau_n).$$

It follows from this equation that

$$\tau_{n+1} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j,$$

which by the strong law of large numbers converges with probability 1 to the expectation $E(\mathbf{x})$ provided the random variables have finite variance.