HUT , Institute of mathematics Gripenberg
Mat-1.196 Mathematics of neural networks
Exercise 4
5.2–13.2.2002

1. Construct, using Kolmogorov's theorem, a neural network that calculates the values of the continuous function $f(x,y)$ when $|x| \le 1$ and $|y| \le 1$.
Solution: According to Kolmogorov's theorem we have

$$f(x,y) = \sum_{k=1}^{5} g\left(\lambda_1\phi_k(x) + \lambda_2\phi_k(y)\right).$$

We construct a net with dimensions $(2,10,5,1)$ such that all thresholds τ are 0, $W_1(i,j) = \frac{1}{2}(1-(-1)^{i+j})$, $W_2(i,j) = \lambda_1\delta_{2i-1,j} + \lambda_2\delta_{2i,j}$, $W_3(i,j) = 1$, $\sigma_1(\mathbf{x})(i) = \sigma_{\lfloor\frac{i+1}{2}\rfloor}(\mathbf{x}(i))$, $\sigma_2(\mathbf{x})(i) = g(\mathbf{x}(i))$, and $\sigma_3(\underline{t}) = \underline{t}$.

In this net the nodefunctions are thus not the same at each node on the same level.

---

2. Construct, using Kolmogorov's theorem, a neural network with two hidden layers such that the node function is the same at all nodes in each layer, such that the network calculates the values of the continuous function $f(x,y)$ when $|x| \le 1$ and $|y| \le 1$.
Solution: According to Kolmogorov's theorem we have

$$f(x,y) = \sum_{k=1}^{5} g\left(\lambda_1\phi_k(x) + \lambda_2\phi_k(y)\right).$$

First we define the numbers c_j, $j = 1,\ldots,5$ by $c_1 = 0$, and $c_j = -\phi_j(1)+\phi_{j-1}(0)+c_{j-1}$ for $j = 2,\ldots,5$. If we now define the function ϕ by $\phi(x-j+1) = \phi_j(x)+c_j$ when $x \in [0,1]$ and $j = 1,\ldots,5$, then ϕ is a continuous and monotone function.

Now we construct a net with dimensions $(2,10,5,1)$ such that $W_1(i,j) = \frac{1}{2}(1-(-1)^{i+j})$, $W_2(i,j) = \lambda_1\delta_{2i-1,j} + \lambda_2\delta_{2i,j}$, $W_3(i,j) = 1$, $\tau_1(i) = \lfloor\frac{i+1}{2}\rfloor - 1$, $\tau_2(i) = (\lambda_1+\lambda_2)c_i$, $\tau_3 = 0$, $\sigma_1 = \phi$, $\sigma_2 = g$ and $\sigma_3(\underline{t}) = \underline{t}$.

---

3. Let B be a symmetric $d\times d$ matrix. Find the point $\mathbf{x}_1$ where the function $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}B\mathbf{x}$ ($\mathbf{x}$ is a $d\times 1$ column vector) gets its smallest value on the line $\mathbf{x}_0 + t\mathbf{u}$, $t \in \mathbb{R}$ where $\mathbf{u} \ne \mathbf{0}$. Calculate the answer $\mathbf{x}_1 = (x_1, y_1)^{\mathrm{T}}$ when $B = \begin{pmatrix}\lambda & 0\\ 0 & 1\end{pmatrix}$, $\mathbf{x}_0 = (x_0, y_0)$ and $\mathbf{u} = f'(\mathbf{x}_0)$. Show that if $|\lambda x_0| = |y_0|$, then $|\lambda x_1| = |y_1|$. What can be said about $\left|\frac{y_1}{y_0}\right|$ in this case.

Solution: Let $g(t) = f(\mathbf{x}_0+t\mathbf{u}) = (\mathbf{x}_0+t\mathbf{u})^{\mathrm{T}}B(\mathbf{x}_0+t\mathbf{u})$. Then $g'(t) = 2\mathbf{u}^{\mathrm{T}}B(\mathbf{x}_0+t\mathbf{u})$ and if we solve the equatin $g'(t) = 0$ we get

$$t = -\frac{\mathbf{u}^{\mathrm{T}}B\mathbf{x}_0}{\mathbf{u}^{\mathrm{T}}B\mathbf{u}},$$

and

$$\mathbf{x}_1 = \mathbf{x}_0 - \frac{\mathbf{u}^{\mathrm{T}}B\mathbf{x}_0}{\mathbf{u}^{\mathrm{T}}B\mathbf{u}}\mathbf{u}.$$

If now $B = \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}$ $\mathbf{x}_0 = (x_0, y_0)^{\mathrm{T}}$ and $\mathbf{u} = f'(\mathbf{x}_0) = 2B\mathbf{x}_0 = 2\begin{pmatrix} \lambda x_0 \\ y_0 \end{pmatrix}$, then

$$\mathbf{x}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} - \frac{\lambda^2 x_0^2 + y_0^2}{\lambda^3 x_0^2 + y_0^2} \begin{pmatrix} \lambda x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} \dfrac{y_0^2 - \lambda y_0^2}{\lambda^3 x_0^2 + y_0^2} x_0 \\ \dfrac{\lambda^3 x_0^2 - \lambda^2 x_0^2}{\lambda^3 x_0^2 + y_0^2} y_0 \end{pmatrix}.$$

If now $|\lambda x_0| = |y_0|$ then we have

$$\frac{|\lambda x_1|}{|y_1|} = \frac{|y_0^2(1-\lambda)\lambda x_0|}{|\lambda^2 x_0^2(\lambda - 1) y_0|} = 1.$$

Furthermore, we have in this case

$$\left|\frac{y_1}{y_0}\right| = \frac{|\lambda^2 x_0^2(1-\lambda)|}{|\lambda\lambda^2 x_0^2 + y_0^2|} = \left|\frac{\lambda - 1}{\lambda + 1}\right|.$$

If either λ is very large or λ is very small, then $\left|\frac{y_1}{y_0}\right|$ is close to 1, that is, a large number of steps are required before y_n gets close to 0, which is the point where the minimum of f is achieved.