HUT , Institute of mathematics — Gripenberg

Mat-1.196 Mathematics of neural networks

Exercise 5

12.2–20.2.2002

1. Show that if in the conjugate gradient method one chooses the new direction to be

$$\mathbf{s}_{k+1} = -f'(\mathbf{x}_{k+1})^{\mathrm{T}} + \gamma_k \mathbf{s}_k,$$

where

$$\gamma_k = \frac{(f'(\mathbf{x}_{k+1}) - f'(\mathbf{x}_k)) \cdot f'(\mathbf{x}_{k+1})}{f'(\mathbf{x}_k) \cdot f'(\mathbf{x}_k)},$$

then one gets the same sequence of points for quadratic functions as for the standard conjugate gradient method.

Solution: If the points $\mathbf{x}_k$ are constructed accroding to the standard conjugate gradient method, then one of their properties in the case of a quadratic function is that $f'(\mathbf{x}_k) \cdot f'(\mathbf{x}_{k+1}) = 0$. But then

$$\frac{(f'(\mathbf{x}_{k+1}) - f'(\mathbf{x}_k)) \cdot f'(\mathbf{x}_{k+1})}{f'(\mathbf{x}_k) \cdot f'(\mathbf{x}_k)} = \frac{f'(\mathbf{x}_{k+1}) \cdot f'(\mathbf{x}_{k+1})}{f'(\mathbf{x}_k) \cdot f'(\mathbf{x}_k)},$$

that is, in this case the method is actually the standard conjugate gradient method.

---

2. Show that if $0 < \rho \leq \sigma < 1$ and h is a continuously differentiable function on $\mathbb{R}$ such that h is bounded from below and $h'(0) < 0$, then there is a number $t > 0$ such that

$$h(t) \leq h(0) + t\rho h'(0),$$
$$|h'(t)| \leq -\sigma h'(0).$$

Solution: First we observe that since $0 < \rho < 1$ and h' is continuous, there exists a number $\delta > 0$ such that if $|s| < \delta$ then $h'(s) < \rho h'(0)$. It follows by the mean value theorem that if $0 < \tau \leq \delta$ then $h(\tau) < h(0) + \tau\rho h'(0)$. Now we let

$$t_0 = \inf\{\, \tau > 0 \mid h(t) = h(0) + \tau\rho h'(0)\,\}.$$

Since $h'(0) < 0$ and h is bounded from below, we know that $t_0 \geq 0$ and by the argument above we know that $t_0 \geq \delta$. If $h'(t_0) > 0$ there must, since $h'(0) < 0$, be some point $t \in (0, t_0)$ such that $h'(t) = 0$ and this point satisfies the desired properties by the definition of t_0. If $h'(t_0) \leq 0$ we choose $t = t_0$ and we have only to show that $h'(t_0) \geq -\rho h'(t_0)$. To see that this is the case let $g(\underline{s}) = h(0) + \underline{s}\rho h'(0) - h'(\underline{s})$. By the definition of t_0 we know that $g(s) \geq 0$ when $s \in [0, t_0]$ and since $g(t_0) = 0$ it follows that $g'(t_0) \leq 0$. Thus we conlcude that $h'(t_0) \geq \rho h'(0)$ which implies that $|h'(t_0)| \leq -\sigma h'(0)$ since $h'(t_0) \leq 0$ and $\sigma \geq \rho \geq 0$.

---

3. Suppose that $\sigma \in (0, \frac{1}{2})$ and that one in the conjugate gradient method chooses the new point $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{s}_k$ so that

$$\left| f'(\mathbf{x}_{k+1}) \cdot \mathbf{s}_k \right| \leq -\sigma f'(\mathbf{x}_k) \cdot \mathbf{s}_k.$$

Show that one then has

$$-\sum_{j=0}^{k} \sigma^j \leq \frac{f'(\mathbf{x}_k) \cdot \mathbf{s}_k}{|f'(\mathbf{x}_k)|^2} \leq -2 + \sum_{j=0}^{k} \sigma^j.$$

and

$$f'(\mathbf{x}_k) \cdot \mathbf{s}_k < 0,$$

for all $k \geq 0$ unless $f'(\mathbf{x}_k) = 0$.

Solution: We use induction and since $\mathbf{s}_0 = -f'(\mathbf{x}_0)$ we have $f'(\mathbf{x}_0) \cdot \mathbf{s}_0 < 0$ unless one starts at a critical point and

$$\frac{f'(\mathbf{x}_0) \cdot \mathbf{s}_0}{|f'(\mathbf{x}_0)|^2} = -1,$$

so the claims hold for $k = 0$.

Suppose the claims hold for $k = n$. By the conjugate gradient method we have

$$\mathbf{s}_{n+1} = -f'(\mathbf{x}_{n+1}) + \frac{|f'(\mathbf{x}_{n+1})|^2}{|f'(\mathbf{x}_n)|^2} \mathbf{s}_n.$$

It follows by our assumptions that

$$\begin{aligned} \frac{f'(\mathbf{x}_{n+1}) \cdot \mathbf{s}_{n+1}}{|f'(\mathbf{x}_{n+1})|^2} &= -1 + \frac{f'(\mathbf{x}_{n+1}) \cdot \mathbf{s}_n}{|f'(\mathbf{x}_n)|^2} \leq -1 - \sigma \frac{f'(\mathbf{x}_n) \cdot \mathbf{s}_n}{|f'(\mathbf{x}_n)|^2} \\ &\leq -1 + \sigma \sum_{j=0}^{n} \sigma^j = -2 + 1 + \frac{\sigma(1 - \sigma^{n+1})}{1 - \sigma} = -2 + \frac{1 - \sigma^{n+2}}{1 - \sigma} = -2 + \sum_{j=0}^{n+1} \sigma^j < 0, \end{aligned}$$

where the last inequlaity follows from the assumption $\sigma < \frac{1}{2}$. Similarly, we get

$$\begin{aligned} \frac{f'(\mathbf{x}_{n+1}) \cdot \mathbf{s}_{n+1}}{|f'(\mathbf{x}_{n+1})|^2} &= -1 + \frac{f'(\mathbf{x}_{n+1}) \cdot \mathbf{s}_n}{|f'(\mathbf{x}_n)|^2} \geq -1 + \sigma \frac{f'(\mathbf{x}_n) \cdot \mathbf{s}_n}{|f'(\mathbf{x}_n)|^2} \\ &\geq -1 - \sigma \sum_{j=0}^{n} \sigma^j = -1 - \frac{\sigma(1 - \sigma^{n+1})}{1 - \sigma} = -\frac{1 - \sigma^{n+2}}{1 - \sigma} = -\sum_{j=0}^{n+1} \sigma^j. \end{aligned}$$

Thus the induction argument works and the proof is completed.

C1. Write a Matlab-function `fun` such that `fun(x, W, dim, sigma`$_1$`,sigma`$_2$`,..)` calculates the the output and intermediate results used by the backpropagation algorithm, when all the weights and thresholds are collected in the vector `W`, the information about the network is in the vector `dim` and the nodefunctions are `sigma`$_1$ etc.