Artificial Neural Networks

Lecture 3: Optimal Hebbian Learning

Dr Juan Neirotti

j.p.neirotti@aston.ac.uk

2023

Supervised Learning in Perceptrons

- To understand the principles behind Hebb's update rule for supervised learning.
- To understand and apply the principles of modeling biological system and its statistical treatment.
- To understand the basic mechanisms of optimization.

Supervised Hebbian Learning in Perceptrons

- Hebb algorithm in the on-line scenario. Here each data point is presented only once and then discarded. The data points are of the form $\mathcal{D} = \{(t_\ell, \boldsymbol{x}_\ell)\}_{\ell=1}^{p}$ where $t_\ell \in \{-1, +1\}$ is the correct classification, according to $\boldsymbol{B} \in \{-1, +1\}^N$, $\|\boldsymbol{B}\|_2 := \sqrt{\langle \boldsymbol{B} \,|\boldsymbol{B} \rangle} = \sqrt{N}$, $t_\ell := \operatorname{sgn}(\langle \boldsymbol{B} \,|\boldsymbol{x}_\ell \rangle)$, and $\boldsymbol{x}_\ell \in \{-1, +1\}^N = \mathscr{X}$ is the pattern to be classified. $|\boldsymbol{B}\rangle$ is the *supervisor* or *teacher* that indicates whether an example is correctly classified or not.
- Settings: $\mathcal{D} = \{(\boldsymbol{x}_\ell, t_\ell)\}_{\ell=1}^{p}$.

$$|\boldsymbol{w}_{\ell+1}\rangle = |\boldsymbol{w}_\ell\rangle + \eta \frac{t_\ell \,|\boldsymbol{x}_\ell\rangle}{\sqrt{N}}.$$

Supervised Hebbian Learning in Perceptrons

- Our objective is to find the η that will produce the fastest decay of the generalization error

$$\varepsilon_G\left(|\boldsymbol{w}\rangle\right) = \int \mathrm{d}\boldsymbol{x}\mathcal{P}_{\boldsymbol{X}}(\boldsymbol{x})\Theta\left(-\frac{\langle \boldsymbol{B}\,|\boldsymbol{x}\,\rangle\,\langle \boldsymbol{w}\,|\boldsymbol{x}\,\rangle}{N}\right)$$

  per iteration step.
- We assume η is a function of quantities we can estimate during the learning process.

Optimal Hebbian Learning

- Observe that we suppose there exists $\boldsymbol{B} \in \{-1, +1\}^N$ such that $t(\boldsymbol{x}) = \operatorname{sgn}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x})$, thus the solution of the learning process is $\boldsymbol{w}^{\star} = \alpha \boldsymbol{B}$ where $\alpha \in \mathbb{R}_+$ is a positive real number.
- Let as define the following parameters:

$$Q_n = \frac{\langle \boldsymbol{w}_n | \boldsymbol{w}_n \rangle}{N}$$

$$R_n = \frac{\langle \boldsymbol{w}_n | \boldsymbol{B} \rangle}{N\sqrt{Q_n}}$$

the normalized length of $|\boldsymbol{w}_n\rangle$ and the cosine of the angle between $|\boldsymbol{w}_n\rangle$ and $|\boldsymbol{B}\rangle$, respectively.

.

- The stochastic variables of the problem are:

$$h_n = \frac{\langle \boldsymbol{w}_n | \boldsymbol{x}_n \rangle}{\sqrt{NQ_n}}, \qquad \phi_n = t_n h_n$$
$$b_n = \frac{\langle \boldsymbol{B} | \boldsymbol{x}_n \rangle}{\sqrt{N}}, \qquad \beta_n = t_n b_n.$$

Learning Equations

- By using the update rule for $\boldsymbol{w}$ and the definitions of R_n and Q_n we have that, in leading order in $1/N$, :

$$\frac{Q_{n+1} - Q_n}{1/N} = 2\eta\sqrt{Q_n}\phi_n + \eta^2$$

$$\frac{R_{n+1} - R_n}{1/N} = \frac{\eta}{\sqrt{Q_n}}(\beta_n - R_n\phi_n) - \frac{\eta^2}{2}\frac{R_n}{Q_n} + O(N^{-1/2}).$$

- To obtain the equations of evolution for this system we need to take the expectation over the variables ϕ and β in the limit of $N \to \infty$.

Probability distributions

- We can also prove that, by using the properties of $\mathcal{P}_{\boldsymbol{X}}(\boldsymbol{x})$, the joint probability of the variables is:

$$\mathcal{P}_{H,B}(h,b) = \mathcal{N}(h)\mathcal{N}(b|Rh, 1-R^2).$$

- The probability of the variables known available to the network is:

$$\mathcal{P}_{T,H}(t,h) = \int \mathrm{d}b\, \Theta(tb)\mathcal{P}_{H,B}(h,b) = 2\mathcal{N}(h)\mathcal{H}\left(-\frac{R\,t\,h}{\sqrt{1-R^2}}\right)$$

where $\mathcal{H}(x) = \int_x^\infty \mathrm{d}y \mathcal{N}(y)$.

- The probability of the variables ϕ and β are then

$$\mathcal{P}_{\Phi}(\phi) = 2\mathcal{N}(\phi)\mathcal{H}\left(-\frac{R\phi}{\sqrt{1-R^2}}\right),$$

$$\mathcal{P}_{B,\Phi}(\beta,\phi) = 2\mathcal{N}(\phi)\mathcal{N}(\beta|R\phi, 1-R^2).$$

Learning Equations

- In the limit of $N \to \infty$

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \lim_{N\to\infty} \int \mathrm{d}\phi_n \mathrm{d}\beta_n \mathcal{P}_{B,\Phi}(\beta_n, \phi_n) \frac{Q_{n+1} - Q_n}{1/N}$$
$$= \int \mathrm{d}\phi \, \mathcal{P}_\Phi(\phi) \left(2\,\eta\,\phi\,\sqrt{Q} + \eta^2\right)$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \lim_{N\to\infty} \int \mathrm{d}\phi_n \mathrm{d}\beta_n \mathcal{P}_{B,\Phi}(\beta_n, \phi_n) \frac{R_{n+1} - R_n}{1/N}$$
$$= \int \mathrm{d}\phi \, \mathcal{P}_\Phi(\phi) \left[\frac{\eta}{\sqrt{Q}} \left(\mathbb{E}_{B|\Phi}[\beta|\phi] - R\phi\right) - \frac{\eta^2}{2}\frac{R}{Q}\right]$$

Optimization

- By minimizing functional variations of η in the equation of motion of R we have that:

$$
\begin{aligned}
\frac{\delta}{\delta\eta(\phi_0)}\frac{\mathrm{d}R}{\mathrm{d}t} &= \lim_{\lambda\to 0}\frac{\mathrm{d}}{\mathrm{d}\lambda}\int \mathrm{d}\phi\,\mathcal{P}_\Phi(\phi)\frac{[\eta+\lambda\delta(\phi-\phi_0)]}{\sqrt{Q}}\left(\mathbb{E}_{B|\Phi}[\beta|\phi]-R\phi\right) \\
&\quad - \frac{R}{2Q}\lim_{\lambda\to 0}\frac{\mathrm{d}}{\mathrm{d}\lambda}\int \mathrm{d}\phi\,\mathcal{P}_\Phi(\phi)\,[\eta+\lambda\delta(\phi-\phi_0)]^2 \\
&= \frac{\mathbb{E}_{B|\Phi}[\beta|\phi_0]-R\phi_0}{\sqrt{Q}} - \eta\frac{R}{Q} = 0 \\
\eta(\phi_0) &= \frac{\sqrt{Q}}{R}\left(\mathbb{E}_{B|\Phi}[\beta|\phi_0]-R\phi_0\right).
\end{aligned}
$$

Optimization

- Where

$$\mathbb{E}_{B|\Phi}[\beta|\phi] - R\phi = \sqrt{\frac{1-R^2}{2\pi}} \frac{\exp\left(-\frac{R^2\phi^2}{2(1-R^2)}\right)}{\mathcal{H}\left(-\frac{R\phi}{\sqrt{1-R^2}}\right)}$$

Conclusion

- By using the expression of the optimal $\eta(\phi)$ we have that

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = 2\sqrt{Q}\mathbb{E}_{\Phi}[\eta(\phi)\phi] + \mathbb{E}_{\Phi}[\eta^2(\phi)] = \mathbb{E}_{\Phi}[\eta^2(\phi)]$$
$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{R}{2Q}\mathbb{E}_{\Phi}[\eta^2(\phi)],$$

therefore, if the initial conditions are such that $R^2(0) = Q(0)$ then $R(t) = \sqrt{Q(t)}$ for all t and the dynamic of the system is ruled by $\dot{Q} = \mathbb{E}_{\Phi}[\eta^2(\phi)]$ with

$$\eta(\phi) = \sqrt{\frac{1-Q}{2\pi}}\left[\mathcal{H}\left(-\sqrt{\frac{Q}{1-Q}}\phi\right)\right]^{-1}\exp\left(-\frac{Q\phi^2}{2(1-Q)}\right).$$