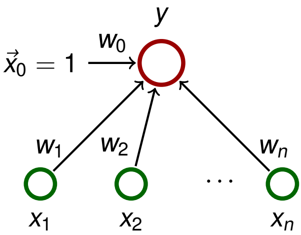
ADALINE

Organizační dynamika:



$\vec{w} = (w_0, w_1, \ldots, w_n)$ a $\vec{x} = (x_0, x_1, \ldots, x_n)$ kde $x_0 = 1$.

Aktivní dynamika:

funkce sítě: $y[\vec{w}](\vec{x}) = \vec{w} \cdot \vec{x}$

ADALINE

Adaptivní dynamika:

- Dána množina **tréninkových vzorů**

$$\mathcal{T} = \left\{ \left(\vec{x}_1, d_1 \right), \left(\vec{x}_2, d_2 \right), \ldots, \left(\vec{x}_p, d_p \right) \right\}$$

  Zde $\vec{x}_k = (x_{k0}, x_{k1} \ldots, x_{kn}) \in \mathbb{R}^{n+1}$, $x_{k0} = 1$, je vstup k-tého vzoru a $d_k \in \mathbb{R}$ je očekávaný výstup.

  Intuice: chceme, aby síť počítala afinní aproximaci funkce, jejíž (některé) hodnoty nám předepisuje tréninková množina.

- **Chybová funkce:**

$$E(\vec{w}) = \frac{1}{2} \sum_{k=1}^{p} \left(\vec{w} \cdot \vec{x}_k - d_k \right)^2 = \frac{1}{2} \sum_{k=1}^{p} \left(\sum_{i=0}^{n} w_i x_{ki} - d_k \right)^2$$

- Cílem je nalézt $\vec{w}$, které minimalizuje $E(\vec{w})$.

Dimenze $n = 1$

Dále budeme uvažovat pouze $n = 1$.

Hodnota sítě pro daný vstup $(1, x_1)$ bude $w_0 + w_1 x_1$

Tedy množina **tréninkových vzorů**

$$\mathcal{T} = \left\{\left(\vec{x}_1, d_1\right), \left(\vec{x}_2, d_2\right), \ldots, \left(\vec{x}_p, d_p\right)\right\}$$

splňuje $\vec{x}_k = (1, x_{k1}) \in \mathbb{R}^2$ a $d_k \in \mathbb{R}$

Zjednodušíme si notaci a **budeme předpokládat**

$$\mathcal{T} = \left\{(x_1, d_1), \ldots, (x_p, d_p)\right\}$$

kde $x_k \in \mathbb{R}$ a $d_k \in \mathbb{R}$ pro $k = 1, \ldots, p$.

Hodnota sítě s váhami w_0, w_1 pro k-tý vzor bude $w_0 + w_1 x_k$.

Chybová funkce pro $n = 1$

$$E(w_0, w_1) = \frac{1}{2} \sum_{k=1}^{p} (w_0 + w_1 x_k - d_k)^2$$

Minimalizujeme E vzhledem k w_0 a w_1:

$$\frac{\delta E}{\delta w_0} = 0 \quad \Leftrightarrow \quad w_0 = \bar{d} - w_1 \bar{x} \quad \Leftrightarrow \quad \bar{d} = w_0 + w_1 \bar{x}$$

kde $\bar{x} = \frac{1}{p} \sum_{k=1}^{p} x_k$ a $\bar{d} = \frac{1}{p} \sum_{k=1}^{p} d_k$

$$\frac{\delta E}{\delta w_1} = 0 \quad \Leftrightarrow \quad w_1 = \frac{\frac{1}{p} \sum_{k=1}^{p} (d_k - \bar{d})(x_k - \bar{x})}{\frac{1}{p} \sum_{k=1}^{p} (x_k - \bar{x})^2}$$

(tj. $w_1 = cov(d, x)/var(x)$)

Normální rozdělení pravděpodobnosti

Rozdělení spojité náhodné veličiny (tj. s hodnotami v $\mathbb{R}$)

Hustota pravděpodobnosti

$$f(x) \quad = \quad \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad =: \quad N[\mu,\sigma^2](x)$$

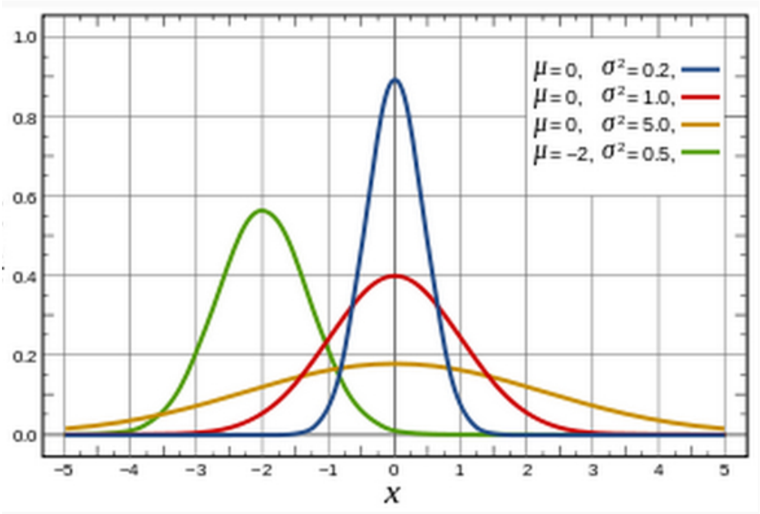
μ je střední hodnota, σ^2 rozptyl

Pokud má náhodná veličina X normální rozdělení, pak

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} p(x)$$

Často se používá k vyjádření náhodné chyby, např. chyby měření, způsobené velkým počtem neznámých a vzájemně nezávislých příčin.

Normální rozdělení pravděpodobnosti



Věrohodnost (likelihood)

Fixujme $\mathcal{T} = \left\{(x_1, d_1), (x_2, d_2), \ldots, \left(x_p, d_p\right)\right\}$

Předpokládejme, že d_k bylo vygenerováno *náhodně* takto

$$d_k = w_0 + w_1 x_k + \epsilon_k$$

Zde

- w_0, w_1 jsou **neznámé konstanty**
- ϵ_k jsou generována náhodně s hustotou pravděpodobnosti $N[0, \sigma^2]$ kde σ^2 je **neznámý rozptyl**

Snadno se ukáže, že hustota pravděpodobnosti, se kterou je vygenerováno d_k splňuje

$$p(d_k \mid w_0, w_1, \sigma^2) = N[w_0 + w_1 x_k, \sigma^2](d_k)$$

Předpokládejme, že pro fixní w_0, w_1, σ^2 jsou $\epsilon_1, \ldots, \epsilon_p$ generována **nezávisle**. Pak hustota pravděpodobnosti, se kterou jsou vygenerována všechna $d_1, \ldots, d_p$ splňuje

$$p(d_1, \ldots, d_p \mid w_0, w_1, \sigma^2) = \prod_{k=1}^{p} N[w_0 + w_1 x_k, \sigma^2](d_k)$$

Maximální věrohodnost (maximum likelihood)

Chceme nalézt w_0, w_1, σ^2, která maximalizují

$$L(w_0, w_1, \sigma^2) := p(d_1, \ldots, d_p \mid w_0, w_1, \sigma^2)$$

Z technických důvodů budeme raději maximalizovat

$$\log(L(w_0, w_1, \sigma^2))$$

kde $\log(y)$ je přirozený logaritmus, tedy funkce inverzní k e^x.

Zřejmě

$$w_0, w_1, \sigma^2 \text{ maximalizují } \quad L(w_0, w_1, \sigma^2)$$
$$\Leftrightarrow$$
$$w_0, w_1, \sigma^2 \text{ maximalizují } \quad \log(L(w_0, w_1, \sigma^2))$$

Maximální log-věrohodnost (log-likelihood)

Ukážeme, že

$$\log(L(w_0, w_1, \sigma^2)) \quad = \quad -\frac{p}{2}\log 2\pi - p\log\sigma - \frac{1}{2\sigma^2}\sum_{k=1}^{p}(d_k - w_0 - w_1 x_k)^2$$

a tedy pro každé σ^2

$$w_0, w_1 \text{ maximalizují } \quad L(w_0, w_1, \sigma^2)$$
$$\Leftrightarrow$$
$$w_0, w_1 \text{ maximalizují } \quad \log(L(w_0, w_1, \sigma^2))$$
$$\Leftrightarrow$$
$$w_0, w_1 \text{ minimalizují } \quad E(w_0, w_1)$$

Tj. maximalizující w_0, w_1 nezávisí na σ^2.

Maximalizujeme-li vzhledem k σ^2, dostaneme

$$\sigma^2 = \frac{1}{p}\sum_{k=1}^{p}(d_k - w_0 - w_1 x_k)^2$$

(tj. průměrná čtvercová odchylka od žádaných hodnot d_k, jak se dalo čekat)

Věrohodnost (likelihood) - libovolná dimenze dat

Fixujme

$$\mathcal{T} = \left\{\left(\vec{x}_1, d_1\right), \left(\vec{x}_2, d_2\right), \ldots, \left(\vec{x}_p, d_p\right)\right\}$$

kde $\vec{x}_k \in \mathbb{R}^{n+1}$ a $d_k \in \mathbb{R}$ pro $k = 1, \ldots, p$.

Předpokládejme, že d_k bylo vygenerováno *náhodně* takto

$$d_k = \vec{w} \cdot \vec{x}_k + \epsilon_k = \sum_{i=0}^{n} w_k x_{ki} + \epsilon_k$$

Zde

- $\vec{w}$ je vektor **neznámých vah**
- ϵ_k jsou generována náhodně s hustotou pravděpodobnosti $N[0, \sigma^2]$ kde σ^2 je **neznámý rozptyl**

Pro fixní $\vec{w}, \sigma^2$ jsou $\epsilon_1, \ldots, \epsilon_p$ generována **nezávisle**. Pak $d_1, \ldots, d_p$ jsou generována s hustotou

$$p(d_1, \ldots, d_p \mid \vec{w}, \sigma^2) = \prod_{k=1}^{p} N[\vec{w} \cdot \vec{x}_k, \sigma^2](d_k)$$

Maximální log-věrohodnost (log-likelihood)

Pro

$$L(\vec{w},\sigma^2) := p(d_1,\ldots,d_p \mid \vec{w},\sigma^2) = \prod_{k=1}^{p} N[\vec{w}\cdot\vec{x}_k,\sigma^2](d_k)$$

platí

$$\log(L(\vec{w},\sigma^2)) \quad = \quad -\frac{p}{2}\log 2\pi - p\log\sigma - \frac{1}{2\sigma^2}\sum_{k=1}^{p}(d_k - \vec{w}\cdot\vec{x}_k)^2$$

a tedy pro každé σ^2

$$\vec{w} \text{ maximalizuje } \quad L(\vec{w},\sigma^2)$$
$$\Leftrightarrow$$
$$\vec{w} \text{ maximalizuje } \quad \log(L(\vec{w},\sigma^2))$$
$$\Leftrightarrow$$
$$\vec{w} \text{ minimalizuje } \quad E(\vec{w})$$

Tj. maximalizující $\vec{w}$ nezávisí na σ^2.

Max. σ^2 splňuje $\sigma^2 = \frac{1}{p}\sum_{k=1}^{p}(d_k - \vec{w}\cdot\vec{x}_k)^2$