



Vilniaus universitetas
Matematikos ir informatikos fakultetas
Informatikos katedra

Radialinių bazinių funkcijų neuroniniai tinklai

prof. dr. Olga Kurasova
Olga.Kurasova@mii.vu.lt

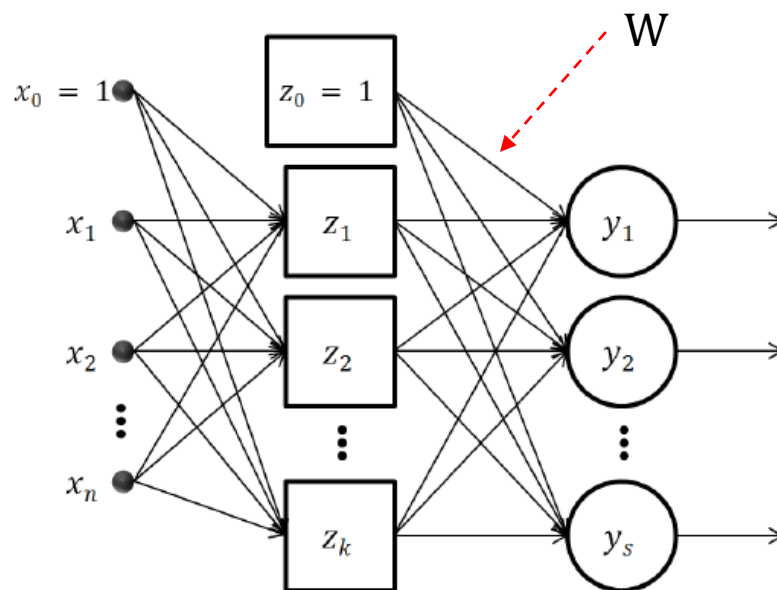
2018

Radialinė bazinė funkcija

- **Radialinė bazinė funkcija** (RBF) – tai funkcija, kurios reikšmės priklauso tik nuo atstumo nuo pradžios taško, t. y. $\phi(X) = \phi(\|X\|)$ arba iki kito taško μ , vadinamojo centro, t. y. $\phi(X, \mu) = \phi(\|X - \mu\|)$.
- Įprastai norma yra **Euklidinis atstumas**, tačiau gali būti naudojami ir kiti atstumai.

Radialinių bazinių funkcijų (RBF) neuroniniai tinklai

- **Radialinių bazinių funkcijų neuroninis tinklas** (angl. *Radial Basis Function Neural Network*, **RBF**) padeda išspręsti funkcijų aproksimavimo, laiko eilučių prognozavimo, klasifikavimo ir kitus duomenų analizės uždavinius.



RBF tinklo sandara

- Tinklas susideda iš n **įėjimų**, vieno **paslėpto** neuronų sluoksnio, kuris sudarytas iš k neuronų ir s **išėjimų**.
- Įėjimo duomenų rinkinys $X = (x_0, x_1, x_2, \dots, x_n)$.
- Paslėptas neuronų sluoksnis $Z = (z_0, z_1, z_2, \dots, z_k)$. Šiame sluoksnyje vietoj aktyvavimo funkcijų yra naudojamos **radialinės bazinės funkcijos**, todėl šis sluoksnis dar yra vadinamas radialinių bazinių funkcijų sluoksniu.
- Išėjimų sluoksnis žymimas $Y = (y_1, y_2, \dots, y_s)$.

Radialinės bazinės funkcijos

- Radialinių bazinių funkcijų **sužadinimo lygis** priklauso nuo atstumo tarp objekto $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$, $i = 1, \dots, m$, ir **radialinės bazinės funkcijos centro** $\mu_j = (\mu_{j1}, \mu_{j2}, \dots, \mu_{jn})$, $j = 1, \dots, k$.
- Bendra radialinių bazinių funkcijų **išraiška** užrašoma taip:

$$z_j = \phi(\|X - \mu_j\|)$$

- Čia $\|X - \mu_j\|$ – atstumas tarp objekto X_i ir centro μ_j , dažniausiai skaičiuojamas **Euklidinis atstumas**, bet gali būti skaičiuojamas ir bet kuris kitas atstumas.

Radialinės bazinės funkcijos

- **Gausinė:**

$$z_j = \exp \left(-\frac{\|X - \mu_j\|^2}{2\sigma^2} \right)$$

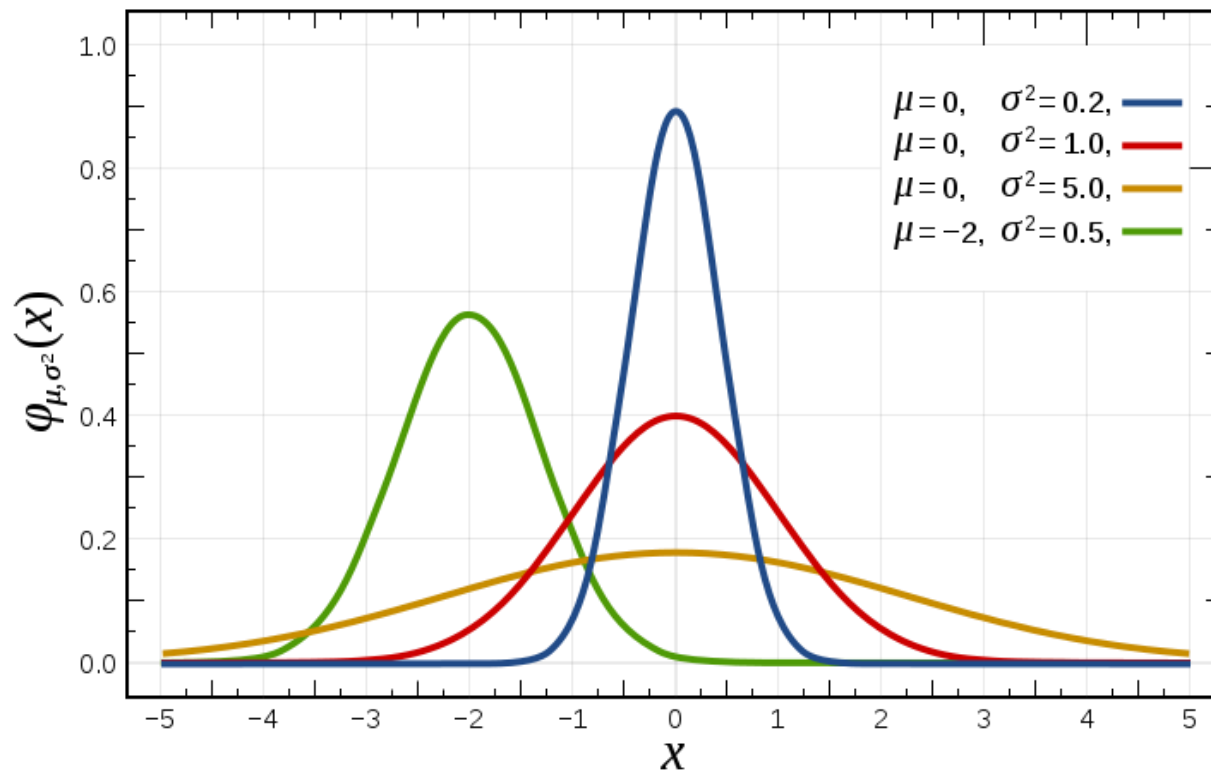
čia σ – pločio parametras.

- **Multikvadratinė:** $z_j = \sqrt{\|X - \mu_j\|^2 + 1}$
- **Multikvadratinė inversija:** $z_j = \frac{1}{\sqrt{\|X - \mu_j\|^2 + 1}}$

Gausinė funkcija

- Dažniausiai RBF neuroniniuose tinkluose yra naudojama **Gausinė** radialinė bazinė funkcija.

$$z_j = \exp\left(-\frac{\|X - \mu_j\|^2}{2\sigma^2}\right)$$



RBF tinklo mokymas

Galimi keli RBF tinklo mokymo būdai:

- **Vieno** etapo,
- **Dviejų** etapų,
- **Trijų** etapų.

Vieno etapo RBF tinklo mokymas

- Mokymo metu nustatomi **antrojo sluoksnio svoriai** W mokymo su mokytoju (*supervised*) būdu.
- RBF **centro taškai** μ_j parenkami iš mokymo duomenų aibės,
- **Parametras** σ prilyginamas iš anksto parinktam skaičiui.

Dviejų etapų RBF tinklo mokymas

- Abu RBF tinklo sluoksniai nustatomi atskirai.
- Pradžioje nustatomi RBF **centro taškai** μ_j ir **parametro** σ reikšmė.
- Vėliau randami **antrojo sluoksnio svoriai** W .
- Tai **dažniausiai** naudojamas būdas.

Trijų etapų RBF tinklo mokymas

- Įvykdomas **dviejų etapų** mokymas.
- Vėliau **visa** tinklo struktūra (μ_j, σ, W) suderinama atlikus tam tikrą **optimizavimo** procedūrą.
- Įprastai optimizavimui reikalingi **pradinės kintamųjų reikšmės**. Tad čia jo prilyginamos rastoms po dviejų etapų mokymo.

RBF tinklo mokymas

- RBF tinklai įprastai **apmokomi dviem etapais**.
- **Pirmame etape** nustatomi RBF parametrai – centro taškai μ_j ir pločio parametras σ .
- Nustačius šiuos parametrus, RBF reikšmės tampa fiksuotos, todėl likusi tinklo dalis yra ekvivalentiška **vienasluoksniam perceptronui**.
- Todėl **antrame etape** tinklo mokymas vyksta minimizuojant paklaidos funkciją gradientiniu nusileidimo algoritmu (keičiami tinklo svoriai).

RBF tinklo mokymas

- RBF centro taškai μ_j gali būti parenkami dvejopai:
- Paprasčiausiu atveju, **pasirenkami atsitiktinai** nagrinėjamos duomenų aibės taškai ir jie laikomi centro taškais.
- Kitas būdas – juos parinkti taikant **k-vidurkių** klasterizavimo algoritmą.
- Yra ir kitų centrų parinkimo būdų.

K-vidurkių algoritmas (1)

- 1) Pasirenkama, **į kiek klasterių** k bus klasterizuojama.
- 2) **Atsitiktinai** parenkami klasterių **centrų taškai**.
- 3) Visi duomenų taškai **priskiriami vienam iš klasterių** pagal jų artumą klasterių centrui.
- 4) Perskaičiuojami **klasterių centrai** ir skaičiuojama kvadratinė paklaida (žr. sekančią skaidrę).
- 5) Jei netenkinama **sustojimo sąlyga**, einama į 3 žingsnį. Priešingu atveju, algoritmas sustabdomas.

K-vidurkių algoritmas (2)

- Vienas iš galimų duomenų panašumo matų yra **kvadratinė paklaida**.
- Tai **Euklido atstumų** tarp kiekvieno klasterio objekto ir klasterio centro kvadratų suma visiems klasteriams:

$$E_{k\text{-means}} = \sum_{j=1}^k \sum_{i=1}^{k_j} \|X_i^j - \mu_j\|^2$$

- Čia k – klasterių skaičius, k_j – j -tajam klasteriui priskirtų duomenų taškų skaičius.
- $\{X_1^j, X_2^j, \dots, X_{k_j}^j\}$ – j -tajam klasteriui priskirtų duomenų taškų aibė.

K-vidurkių algoritmas (3)

K-vidurkių metodo **sustojimo sąlygos**:

- kvadratinės paklaidos reikšmė $E_{k\text{--means}}$ tampa **mažesnė** už pasirinktą slenkstinę reikšmę (norimą mažą skaičių)
- arba duomenų taškai **nebepersiskirsto** kitiems klasteriams.

RBF tinklo mokymas

- Kai centro taškui μ_j ir parametrai σ nustatyti taikomas **k-vidurkių** (k-means) algoritmas, pasirenkama **į kelis klasterius** bus klasterizuojama, priklausomai nuo RBF tinklo struktūros (kiek neuronų pirmame sluoksnyje).
- Visi duomenys k-vidurkių algoritmu **suklasterizuojami** į tiek klasterių. $\{X_1^j, X_2^j, \dots, X_{k_j}^j\}$ – j -tajam klasteriui priskirtų duomenų taškų aibė.
- Gautų **klasterių centrai** tampa centro taškais μ_j .
- Parametras σ gali būti parenkamas kiekvienam klasteriui. Tai gali būti **standartinis nuokrypis**

$$\sigma_j = \frac{1}{k_j} \sum_{i=1}^{k_j} \|X_i^j - \mu_j\|.$$

RBF tinklo mokymas

- Radus centrų taškus ir σ reikšmes, visiems mokymo duomenims **apskaičiuojamos RBF reikšmės**. Gaunama matrica Z .
- Turinti norimų reikšmių vektorių T , **antrojo sluoksnio svoriai** W randami minimizuojant paklaidą:

$$E(W) = \|ZW - T\|^2$$

- Svoriai W gali būti gauti **analitiškai**: $W = Z^+T$, čia Z^+ yra matricos Z pseudo-atvirkštinė matrica, kurios paprastesnė išraiška yra: $Z^+ = (Z^T Z)^{-1} Z^T$.

RBF taikymas duomenims klasifikuoti

- Išėjimų skaičius lygus klasių skaičiui.
- Duomenys suklasterizuojami į pasirinktą k klasterių skaičių.
- Apskaičiuojamos σ reikšmės.
- Visi mokymo duomenys „praleidžiami“ per RBF, pvz., Gausinę funkciją.

RBF taikymas duomenims klasifikuoti

- Tuomet **išėjimų skaičius** lygus klasių skaičiui.
- Duomenys **suklasterizuojami** į pasirinktą k klasterių skaičių, taip randami RBF centro taškai.
- Apskaičiuojamos σ reikšmės (klasterių **standartiniai nuokrypiai**).
- Visi mokymo duomenys „**praleidžiami**“ per RBF, pvz., Gausinę funkciją.
- Randami antrojo sluoksnio **svoriai** W .
- **Nauji duomenys** pateikiami į išmokytą tinklą. Pagal gautas išėjimų reikšmes yra **nustatomos klasės**.
- Patikrinamas **klasifikavimo tikslumas**, t. y., kiek duomenų yra teisingai suklasifikuota.

RBF taikymas funkcijoms aproksimuoti

- Į **įėjimus** bus pateikiamos aproksimuojamos funkcijos kintamųjų reikšmės, norimos **išėjimų reikšmės** bus funkcijos reikšmės.
- Pasirenkamas RBF išėjimų skaičius k .
- Pasirenkama parametro σ reikšmė.
- Taškai, sudaryti iš kintamųjų reikšmių, **suklasterizuojami** į k klasterių.
- Visi taškai „**praleidžiami**“ per RBF, pvz., Gausinę funkciją.
- Randami antrojo sluoksnio **svoriai** W .
- **Taškai** pateikiami į išmokytą tinklą. Pagal gautas išėjimų reikšmes yra **nustatomos funkcijų reikšmės**.

RBF tinklo taikymo pavyzdžiai

- Paprastus RBF tinklo taikymo pavyzdžius galima rasti adresu:
<http://mccormickml.com/2013/08/15/radial-basis-function-network-rbfn-tutorial/>