

Computer Vision 3

Ș.I. dr. ing. Mihai DOGARIU

www.mdogariu.aimultimedialab.ro

Structura cursului



M1. Introducere

M2. Fundamentele Învățării Adânci (Deep Learning Fundamentals)

M3. Învățare Adâncă Supervizată (Supervised Deep Learning)

M4. Învățare Adâncă Nesupervizată (Unsupervised Deep Learning)

M5. Învățare Consolidată (Reinforcement Learning)

M4. Învățare Adâncă Nesupervizată (Unsupervised Deep Learning)

4.1. Concept Unsupervised Deep Learning

4.2. Autoencoders (AE)

4.3. Variational Autoencoders (VAE)

4.4. Generative Adversarial Networks (GAN)

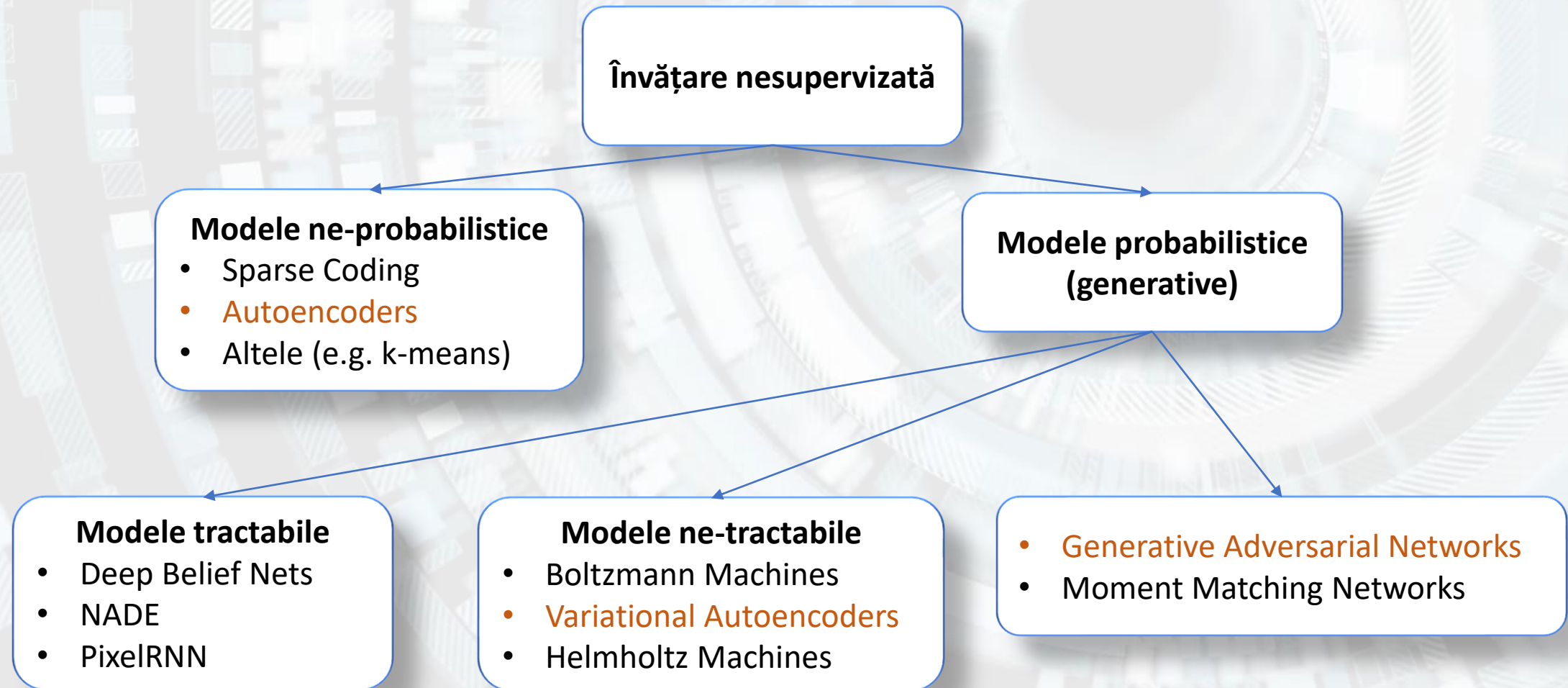
M4.1. Concept Unsupervised Deep Learning

Unsupervised Deep Learning

Învățarea supervizată = paradigmă de învățare a mașinilor în care datele de antrenare sunt etichetate. Fiecare exemplu de antrenare este format dintr-un descriptor de trăsături și o etichetă. Scopul învățării supervizate este de a învăța funcția de asociere dintre trăsăturile de intrare și eticheta corespondentă.

Învățarea nesupervizată = paradigmă de învățare a mașinilor în care datele de antrenare sunt neetichetate. Scopul învățării nesupervizate este de a învăța modelul inherent al exemplelor de antrenare prin imitare. Aceasta se face prin auto-organizare ce captează informații privitoare la comportamentul/distribuția de probabilitate a datelor de antrenare.

Unsupervised Deep Learning



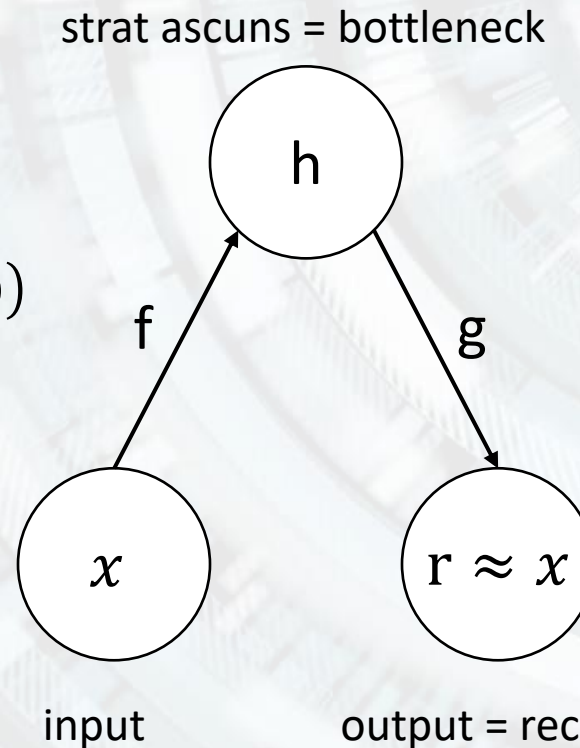
M4.2. Autoencoders

Autoencoders

Autoencoder = rețea neuronală antrenată pentru a încerca să reproducă intrarea la ieșire.

Funcții deterministe
 $encoder: h = f(x)$
 $decoder: r = g(h) = g(f(x))$

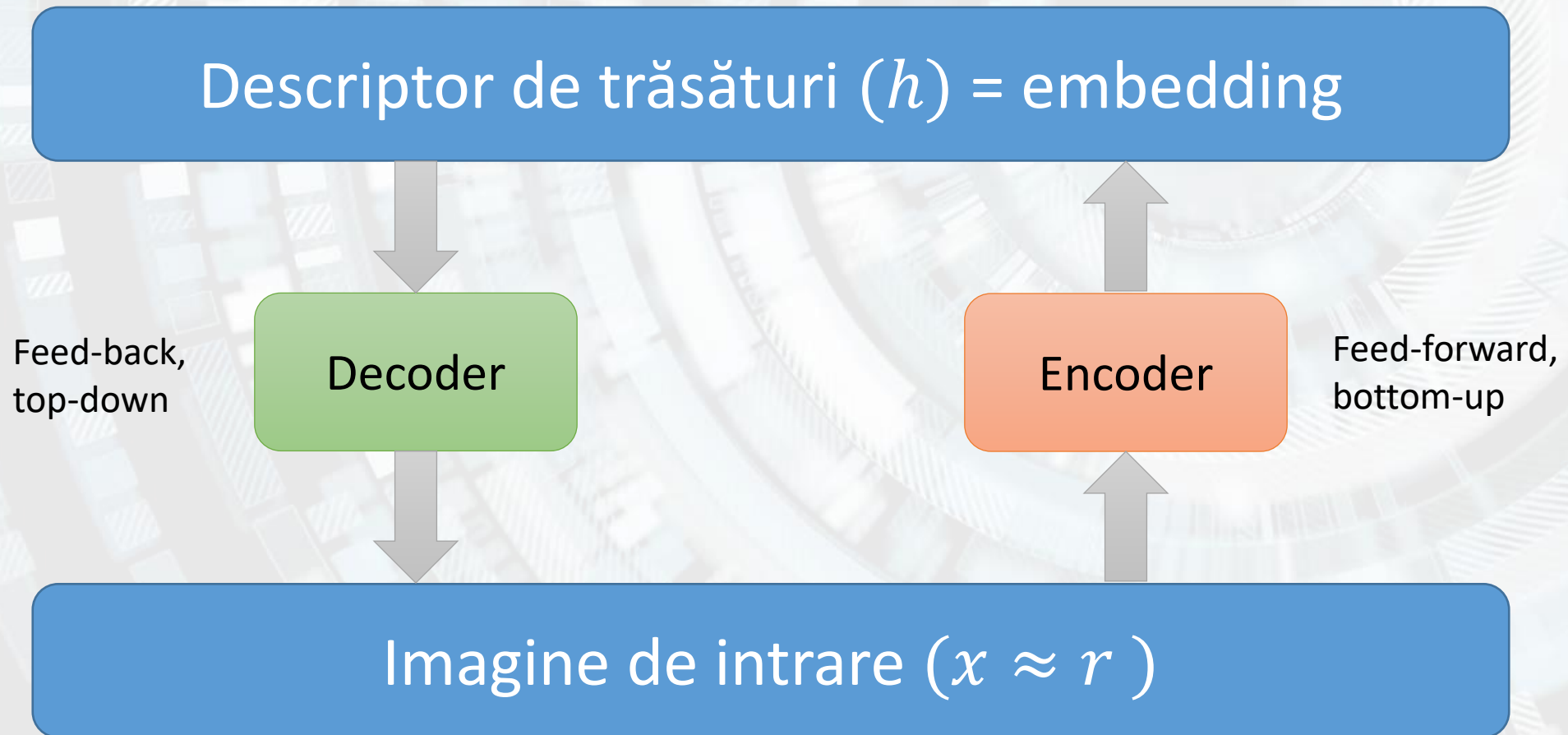
- f și g pot fi obținute cu ajutorul unor rețele neuronale;
- x și r reprezintă structuri informaționale (vectori/tensori);
- h reprezintă un descriptor de trăsături.



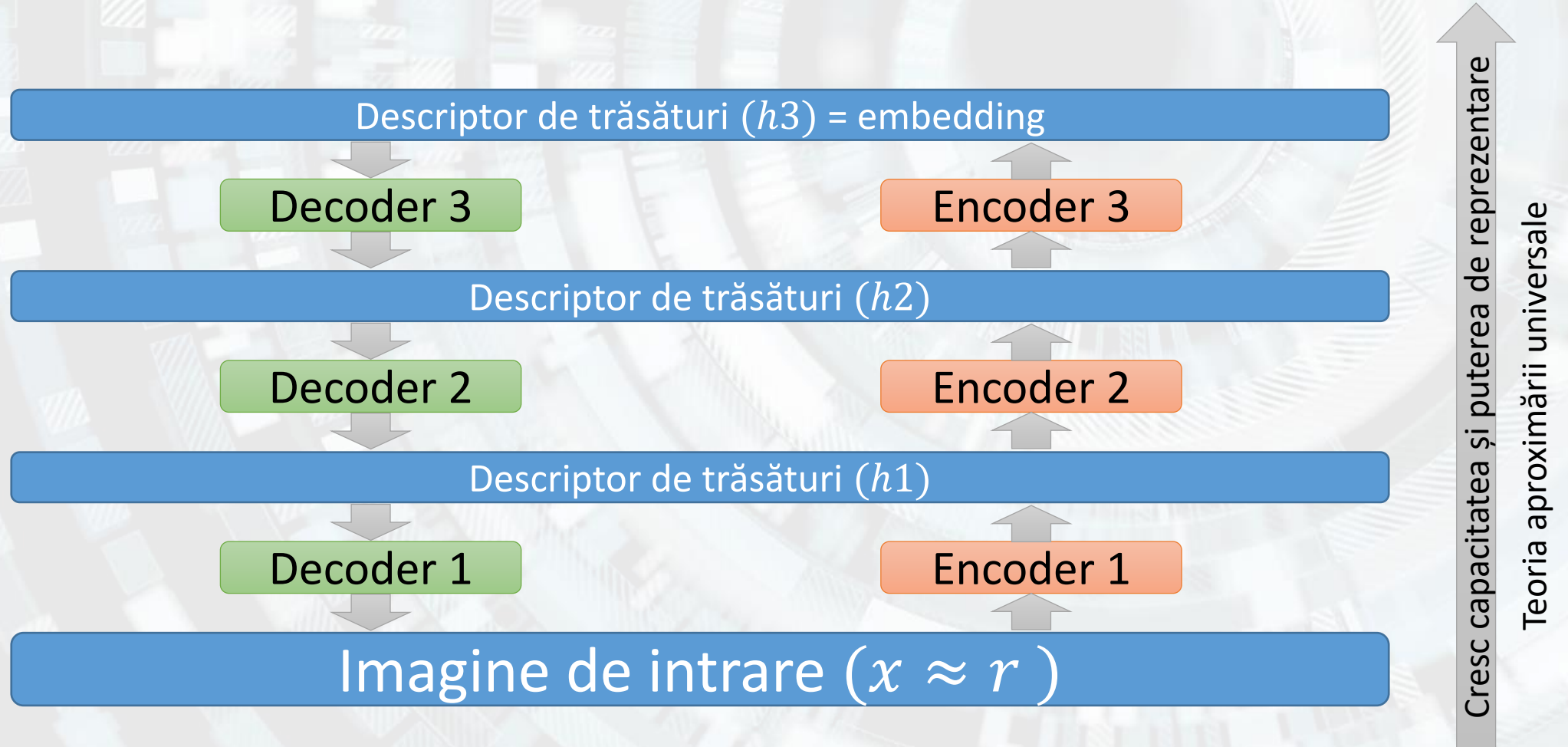
Generalizare stocastică
 $encoder: p_{encoder}(h|x)$
 $decoder: p_{decoder}(x|h)$

Antrenarea se face cu o pierdere
 $L(x, g(f(x)))$
care penalizează ieșirea $g(f(x))$ atunci
când este diferită de x ; e.g. MSE.

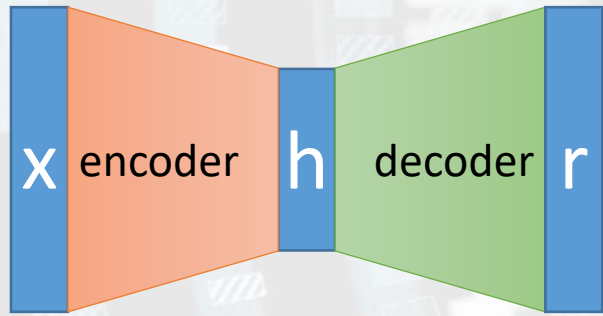
Autoencoders



Autoencoders – stacked



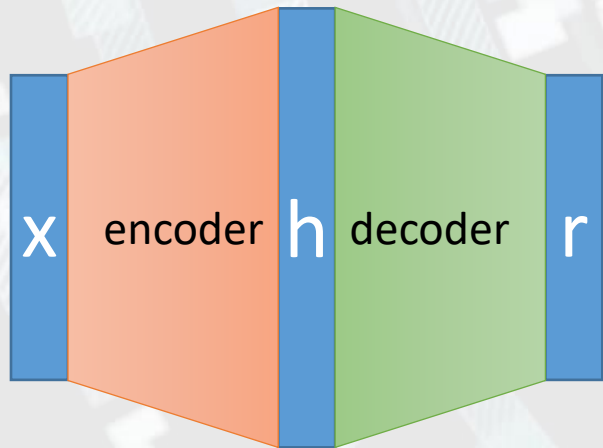
Autoencoders



Undercomplete autoencoder

$$\dim(h) < \dim(x)$$

Encoder-ul și decoder-ul învață o reprezentare contractată a datelor.

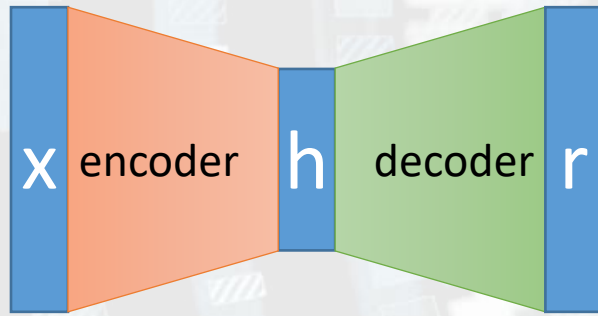


Overcomplete autoencoder

$$\dim(h) \geq \dim(x)$$

Encoder-ul și decoder-ul învață o reprezentare dilatată a datelor.

Autoencoders



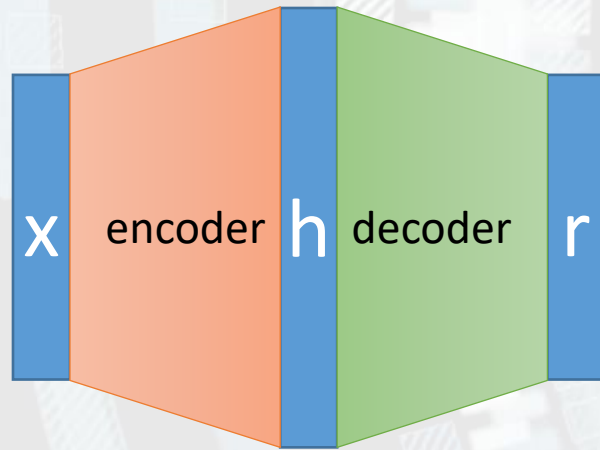
Undercomplete autoencoder

$$\dim(h) < \dim(x)$$

Encoder-ul și decoder-ul învață o reprezentare contractată a datelor.

- Descriptorul de trăsături este restricționat la o dimensiune mai mică decât intrarea => autoencoder-ul învață cele mai reprezentative trăsături;
- Decodor liniar + MSE => autoencoderul învață același subspațiu ca un PCA;
- Codor neliniar + decodor neliniar => reprezentare mai puternică decât PCA;
- Dacă oferim prea multă capacitate (numărul de parametri antrenabili) codorului și decodorului, acestea pot ajunge să învețe funcția identitate => autoencoder-ul devine inutil.

Autoencoders



Overcomplete autoencoder

$$\dim(h) \geq \dim(x)$$

Encoder-ul și decoder-ul învață o reprezentare dilatată a datelor.

- Descriptorul de trăsături are dimensiunea cel puțin egală cu dimensiunea datelor;
- Autoencoder-ul poate reproduce intrarea la ieșire foarte simplu, prin copierea datelor de la un capăt la altul (encoder liniar + decoder liniar) – acest caz nu este dorit, deoarece autoencoder-ul nu învață nimic;
- Funcția identitate este evitată prin regularizare – o serie de constrângeri suplimentare ce împiedică copierea intrării la ieșire. Regularizarea se poate face prin:
 - Reprezentări rare (sparse autoencoder);
 - Introducerea robusteții la zgomot (denoising autoencoder);
 - Păstrarea derivatelor la niște valori mici.

Autoencoders – sparse autoencoder

Se introduce un termen de penalizare a rarității (sparsity penalty) în funcția de cost:

$$\mathcal{L} = L(x, g(f(x))) + \Omega(h)$$

Parametrul de penalizare a rarității este, de obicei, definit ca:

1. Regularizare L1:

$$\Omega(h) = \lambda \sum_i |a_i^{(h)}|$$

Unde $a_i^{(h)}$ este vectorul activărilor din stratul h pentru eșantionul i.

2. Divergență Kullback-Leibler:

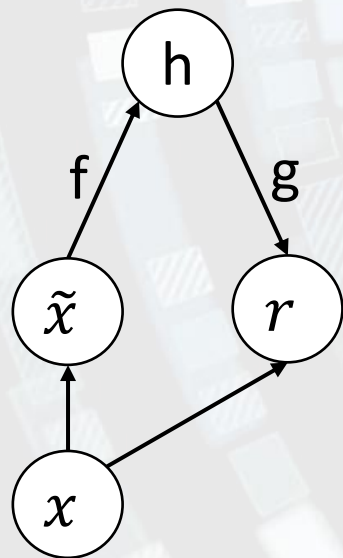
$$\Omega(h) = \sum_j KL(\rho || \hat{\rho}_j)$$

Unde:

- $\hat{\rho}_j = \frac{1}{m} \sum_i |a_i^{(h)}(x)|$ = raritatea neuronului j din stratul h, calculată ca media activărilor a m eșantioane notate cu x;
- ρ este raritatea dorită și reprezintă activarea medie a unui neuron pentru un subset de date.

Autoencoders – denoising autoencoder

Dacă un autoencoder poate reconstrui intrarea, de ce nu ar putea să reconstruiască și o variantă ușor modificată a acesteia?



Intrare: $\tilde{x} = x + n$

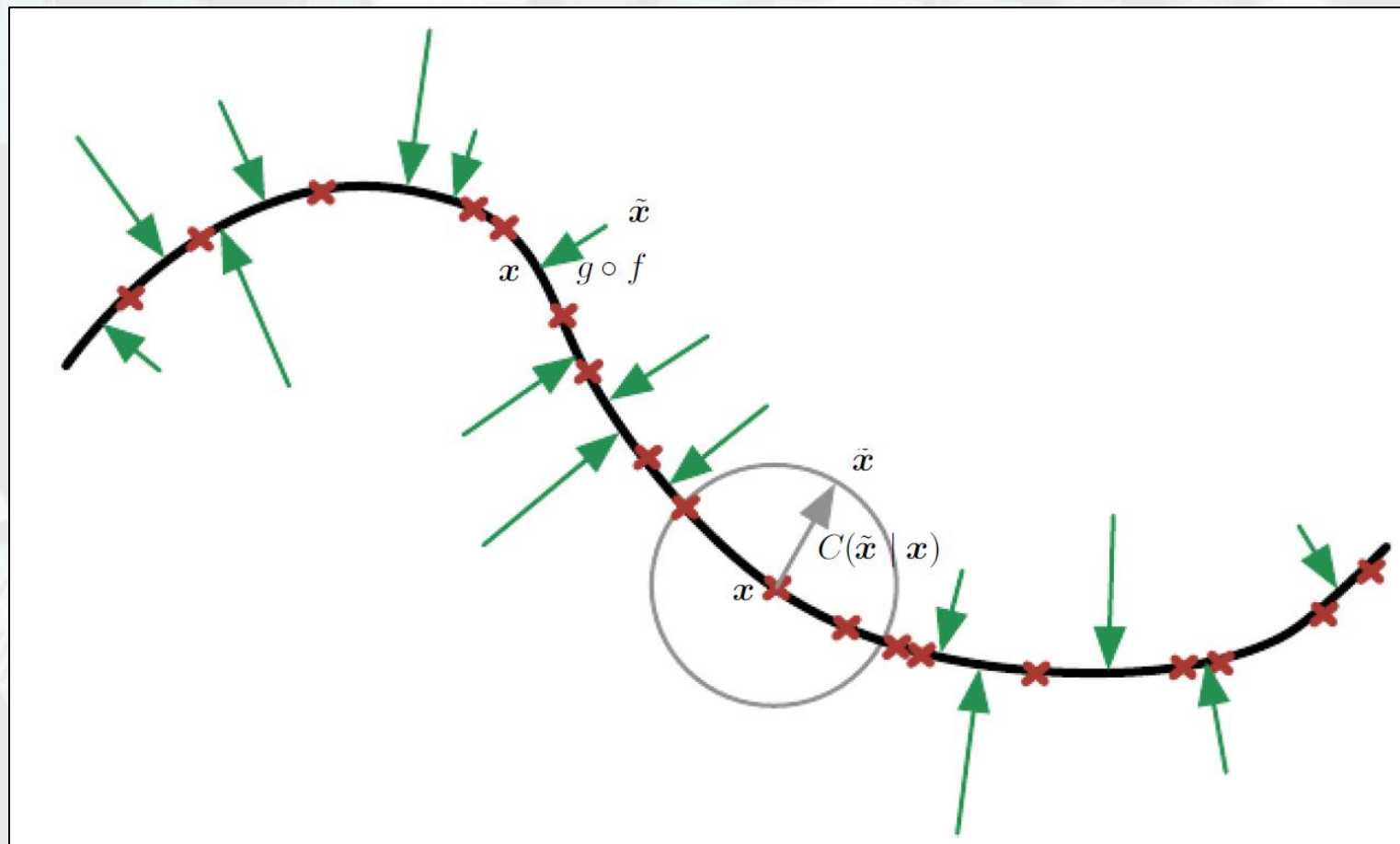
Ieșire (reconstrucție): x

Obiectiv: învățarea modului de eliminare a zgomotului n din intrare

Optimizare: minimizarea funcției de cost

$$\mathcal{L} = L(x, g(f(\tilde{x}))) = L(x, g(f(x + n)))$$

Autoencoders – denoising autoencoder



- Spațiul datelor de antrenare
- × Eșantioane de antrenare
- Coruperea datelor
- Recuperarea datelor

Autoencoders – penalizarea derivatelor

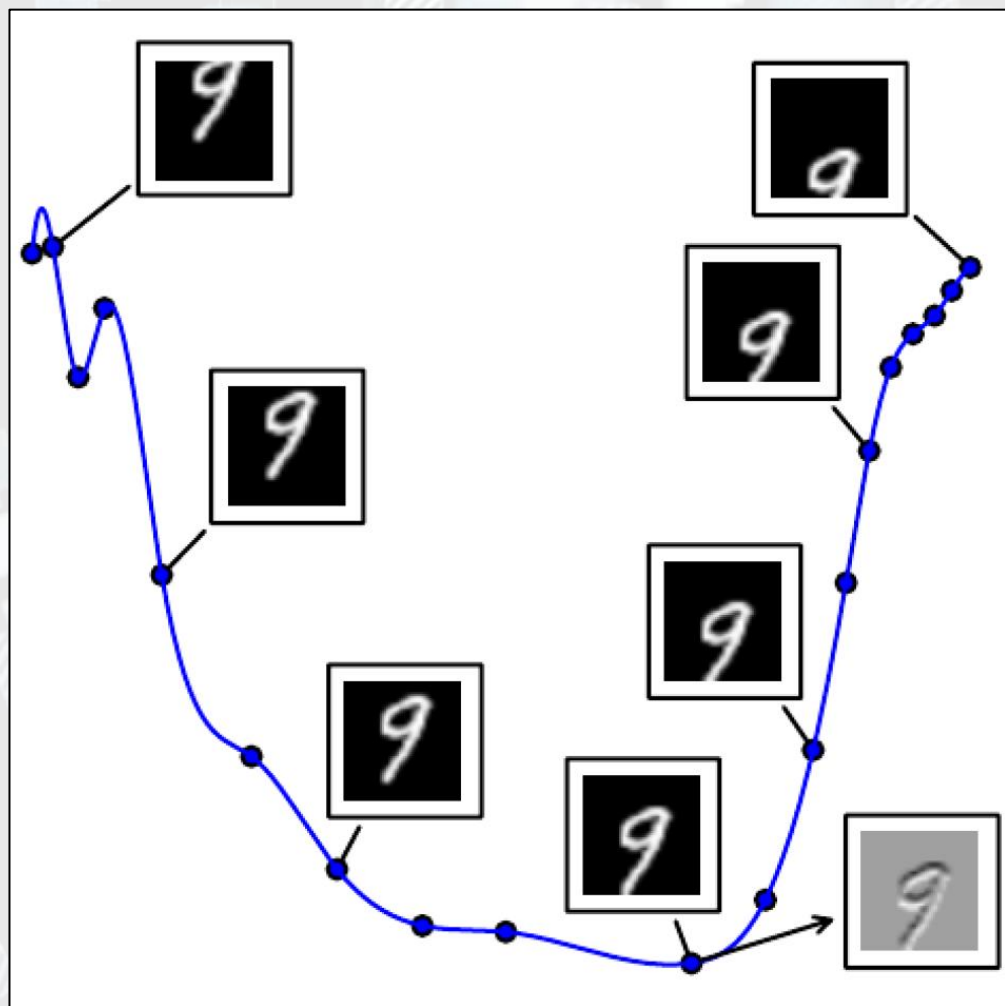
Se introduce un termen de penalizare a rarității (sparsity penalty) în funcția de cost:

$$\mathcal{L} = L(\mathbf{x}, g(f(\mathbf{x}))) + \Omega(\mathbf{h}, \mathbf{x})$$

$$\Omega(\mathbf{h}, \mathbf{x}) = \lambda \sum_i \|\nabla_{\mathbf{x}} h_i\|^2$$

- Forțează modelul să învețe funcții care nu suferă schimbări majore la mici modificări ale lui \mathbf{x}
- Se aplică doar în momentul antrenării, deci forțează modelul să învețe descriptori care captează informații legate de distribuția de probabilitate a datelor de antrenare.
- Se mai numește și autoencoder contractiv, deoarece asociază unei vecinătăți de intrări o vecinătate mai mică de ieșiri.

Autoencoders – învățarea spațiului de date



Asociere între planul imaginii (784-dimensional, imagini din baza MNIST) și curba spațiului de date (manifold) undimensională. În acest caz, s-a folosit reprezentare bidimensională cu ajutorul PCA.

Translația imaginii generează o curbă complexă a spațiului de date.

Hiperplan tangent = plan ce definește direcțiile în care ne putem deplasa cu un pas infinitesimal și să rămânem pe curba spațiului de date. În acest caz, există o singură linie tangentă.

Un spațiu n-dimensional are n linii tangente.

Imaginea gri reprezintă schimbarea suferită în urma avansării în direcția dată de tangentă:

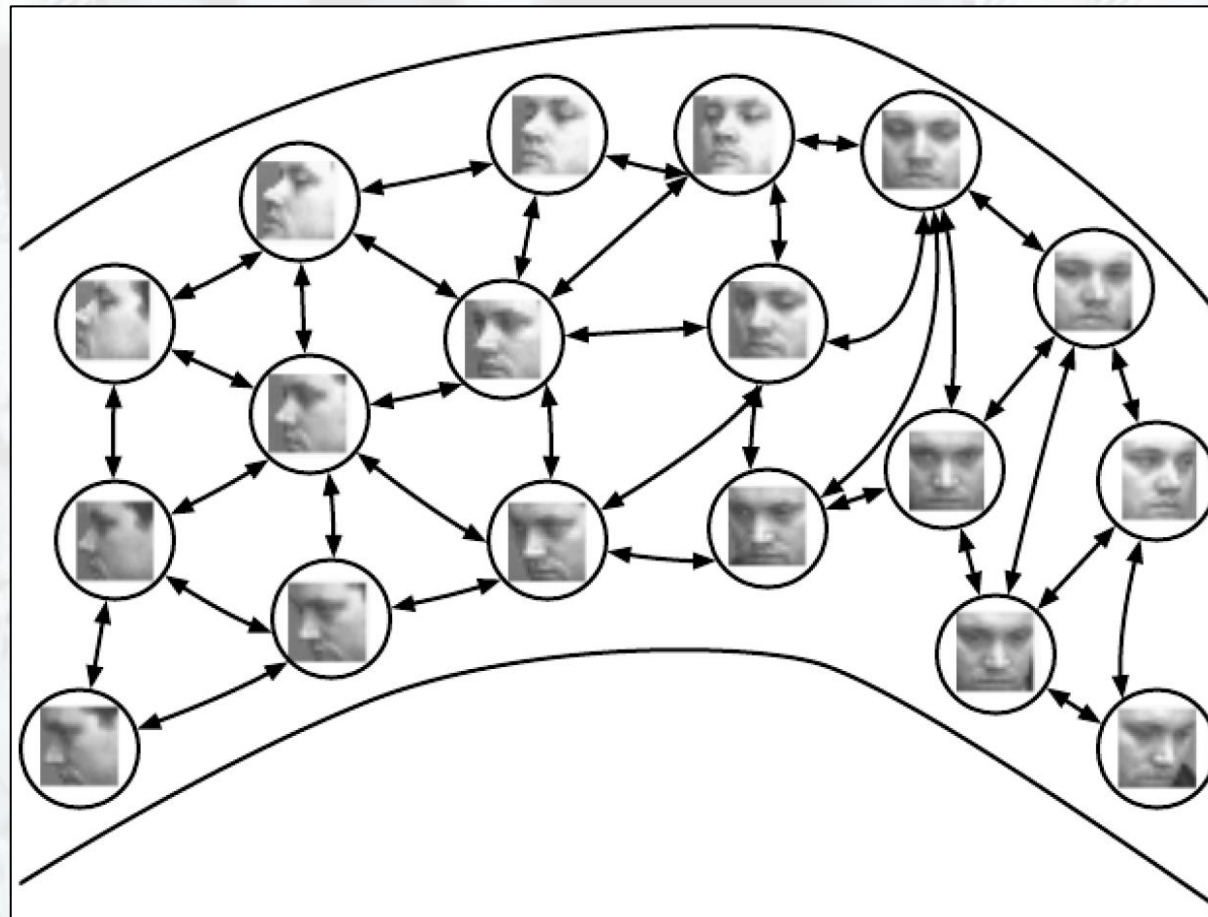
- pixeli gri => pixelii rămân neschimbați;
- pixeli albi => pixelii se deschid la culoare;
- pixeli negri => pixelii se închid la culoare.

Autoencoders – învățarea spațiului de date

Învățarea spațiului de date duce la crearea unui graf cu cei mai apropiați vecini (nearest neighbor):

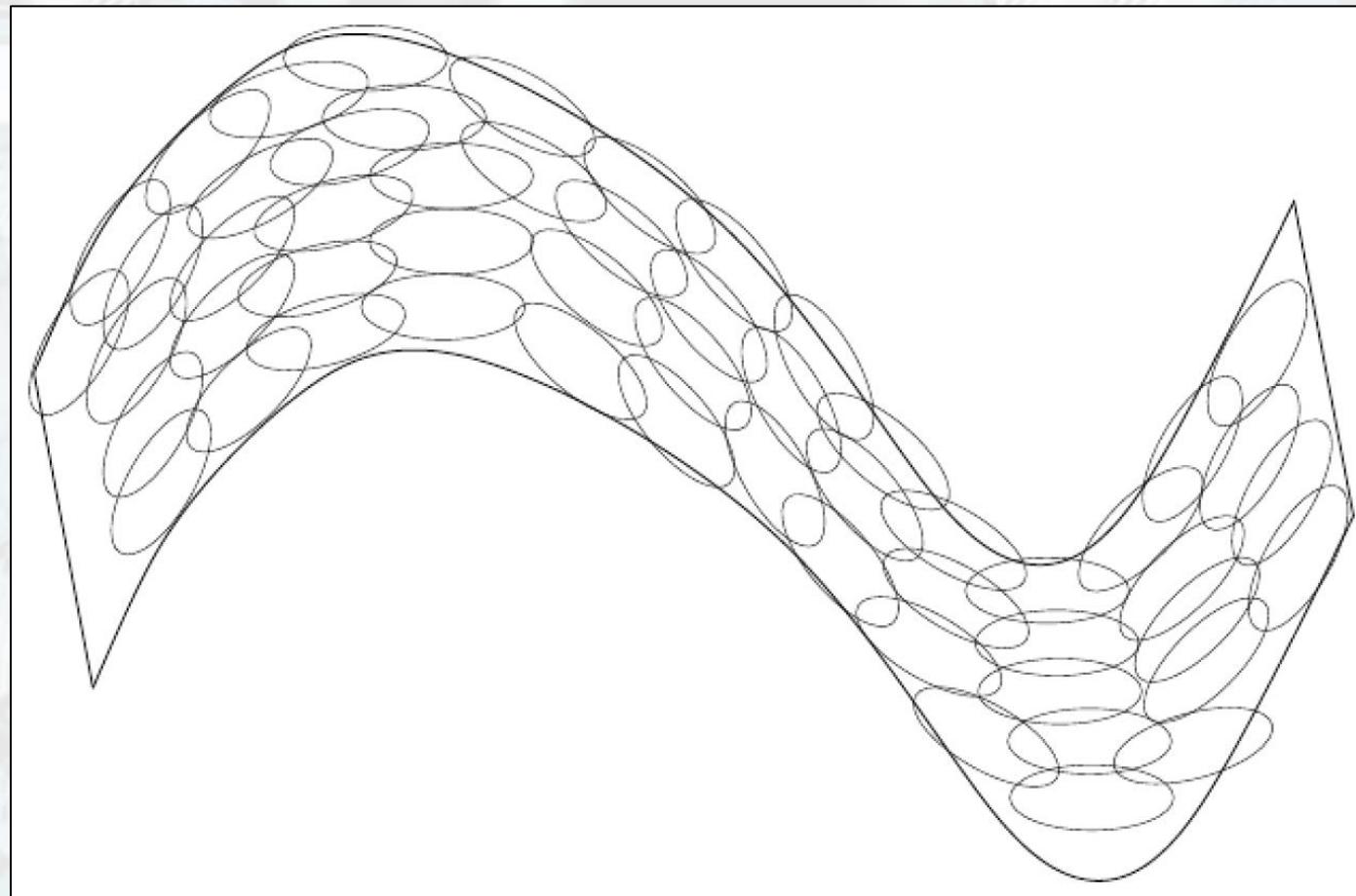
- nodurile = eșantioane de antrenare;
- muchiile = relațiile de tipul cel mai apropiat vecin.

Dacă există suficiente date de antrenare, se poate învăța întregul spațiu de date și, deci, se pot genera eșantioane noi prin interpolare.



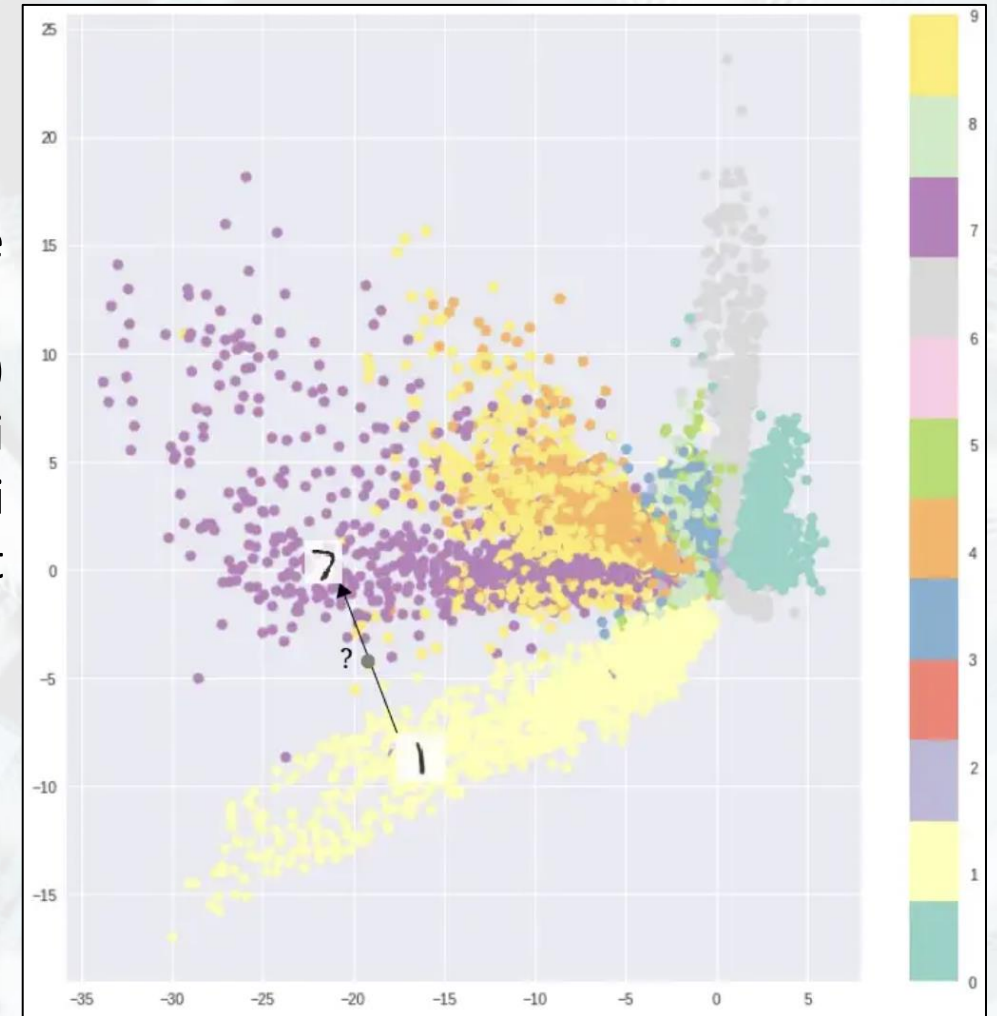
Autoencoders – învățarea spațiului de date

- Cunoașterea planurilor tangente din fiecare locație conduce la combinarea lor și estimarea funcției de densitate a probabilității datelor, care modelează complet setul de date de antrenare.
- Fiecare plan tangent este aproximat cu o curbă Gaussiană.
- În practică, spațiul datelor este foarte complex și poate fi doar aproximat (imposibil de determinat precis).



Autoencoders – învățarea spațiului de date

- Proiectarea descriptorilor într-un spațiu 2D;
- Antrenarea autoencoder-ului pe MNIST Digits;
- Fiecare cifră aparține unui cluster relativ bine definit;
- Alegem un punct ce nu se află pe grafic – ce imagine este decodată?
- **Limitare:** decodorul poate procesa (replica input) doar descriptori pe care i-a văzut în timpul antrenării – spațiu discret de valori. Decodarea altor valori duce la imagini fără sens => nu este adecvat generării de exemple noi.



Autoencoders – concluzii

- Autoencoders sunt foarte bune pentru a recupera dintr-un spațiu de date informația folosită la antrenare;
- Implică antrenarea cu un număr mare de exemple pentru a acoperi întregul spațiu de date (manifold);
- Sunt robuste la o perturbație relativ mică a descriptorului de trăsături (a embedding-ului);
- Nu sunt adecvate generării de exemple noi.

Autoencoders – aplicații

1. **Reducerea dimensionalității datelor** – undercomplete autoencoders pot reduce dimensionalitatea datelor mai mult decât tehnicile clasice, e.g. PCA. O aplicație imediată este compresia datelor în comunicații.
2. **PCA** – în cazul folosirii activărilor liniare, autoencoderele obțin un spațiu latent similar cu PCA.
3. **Information retrieval** – descriptorii din stratul bottleneck pot fi utilizați pentru recuperarea informației (vezi M2, slide 63).
4. **Detecția anomaliilor** – outliers vor fi evidențiați cu ușurință, deoarece vor avea o reconstrucție foarte slabă.
5. **Clustering** – spațiul latent va fi organizat sub forma unor clustere compuse din exemplele cel mai asemănătoare.
6. **Denoising** – autoencoder care primește la intrare imagini cu zgomot și este antrenat să reproducă imagini fără zgomot.
7. **Inpainting** – autoencoder care primește la intrare imagini cu zone ascunse și este antrenat să reproducă imagini complete.
8. **Colorizare** – autoencoder care primește la intrare imagini cu nuanțe de gri și este antrenat să reproducă imagini color.

Sfârșit M4

Bibliografie

- [1] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
- [2] Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.