

Wyjaśnianie predykcji

Uproszczenia modeli

- Funkcja kosztu nie zawsze odpowiada złożoności problemu:
 - Wydajność
 - Etyka
 - Zgodność z prawem
- Środowisko w którym model jest używany jest zmienne (czasem również pod wpływem działania modelu)

Cechy do interpretacji

- Zaufanie
- Przyczynowość
- Transferowalność
- Informatywność
- Sprawiedliwe i etyczne decyzje
- Niezawodność

Interpretowalność modelu

Interpretowalność jest to poziom na którym człowiek jest w stanie zrozumieć przesłanki decyzji

Interpretowalność jest to możliwość przewidywania przez człowieka rezultatów działania modelu

Interpretowalność:

- jak model działa; czy jest zbieżny a jego rozwiązanie jest jednoznaczne
- dlaczego model podał określony wynik

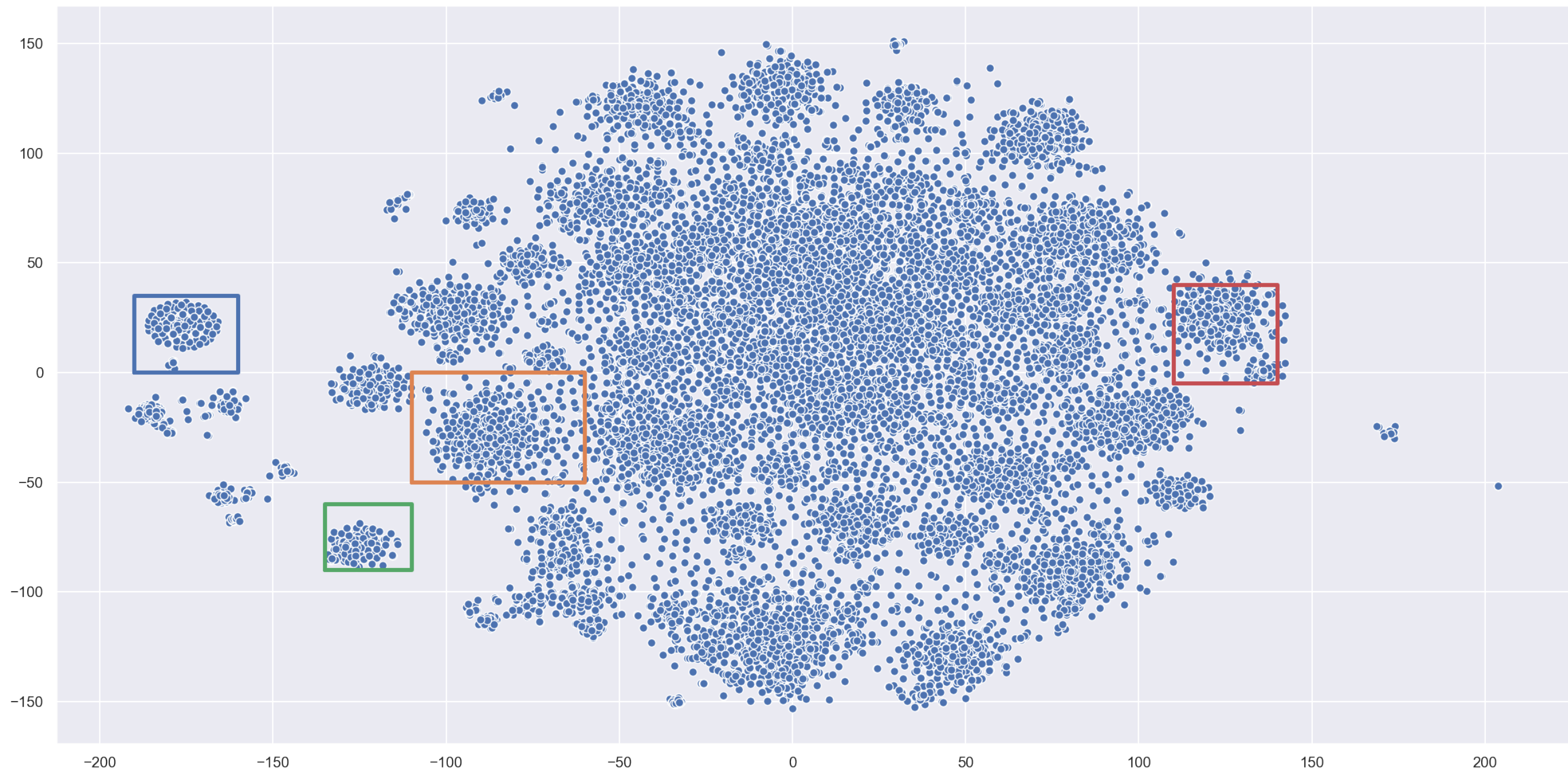
Jak wyjaśnić?

- Z konstrukcji modelu
- Post hoc
 - Tekst
 - Wizualizacja (t-SNE)
 - Wyjaśnienia lokalne - mapa istotności
 - Podać przykład (np. Podobne słowa a w word2vec)
 - Dla sieci neuronowych: cechy wyuczone

Wyjaśnialne modele

Algorithm	Linear	Monotone	Interaction	Task
Linear regression	Yes	Yes	No	regr
Logistic regression	No	Yes	No	class
Decision trees	No	Some	Yes	class,regr
RuleFit	Yes	No	Yes	class,regr
Naive Bayes	No	Yes	No	class
k-nearest neighbors	No	No	No	class,regr

t-SNE



t-SNE

t-Distributed Stochastic Neighbour Embedding

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)},$$

```
from sklearn.manifold import TSNE

unique_tsne_model_pca = TSNE(perplexity=30, n_components=2, init='pca',
                             learning_rate=1500,
                             n_iter=5000, random_state=23)
unique_tsne_embed_pca = unique_tsne_model_pca.fit_transform(uts_pca)
```


Metody niezależne od modelu

- Globalne zastępstwo (surrogate)
- Local interpretable model-agnostic explanations (LIME)
- SHapley Additive exPlanations (SHAP)

LIME dla tekstu

Back

MENU

DASHBOARD

SALES

PATENTS LIST

HOT TOPICS

Controlled-dose dispenser with integral nozzle and cap

Inventors: Steven Lau

Published: 23-12-2003

Abstract

RELATED APPLICATION

The subject application is a continuation of U.S. Ser. No. 09/417,344, which was filed on Oct. 14, 1999, and is entitled "Controlled-Dose Dispenser with Integral Nozzle and Cap".

FIELD OF THE INVENTION

The present invention generally relates to dispensers and, more particularly, to a single-use dispenser for supplying a desired dosage of a liquid substance such as a drug, vaccine, ophthalmic solution and the like in drop form for delivering, including ophthalmic, oral, nasal or ear use or to irrigate a wound or the eyes, for example.

BACKGROUND OF THE INVENTION

A variety of disposable medication dispensers are available on the market. Some are designed for single use, while others are intended for multiple dosage. Some

Classifications

US PO classification

B65D 4710

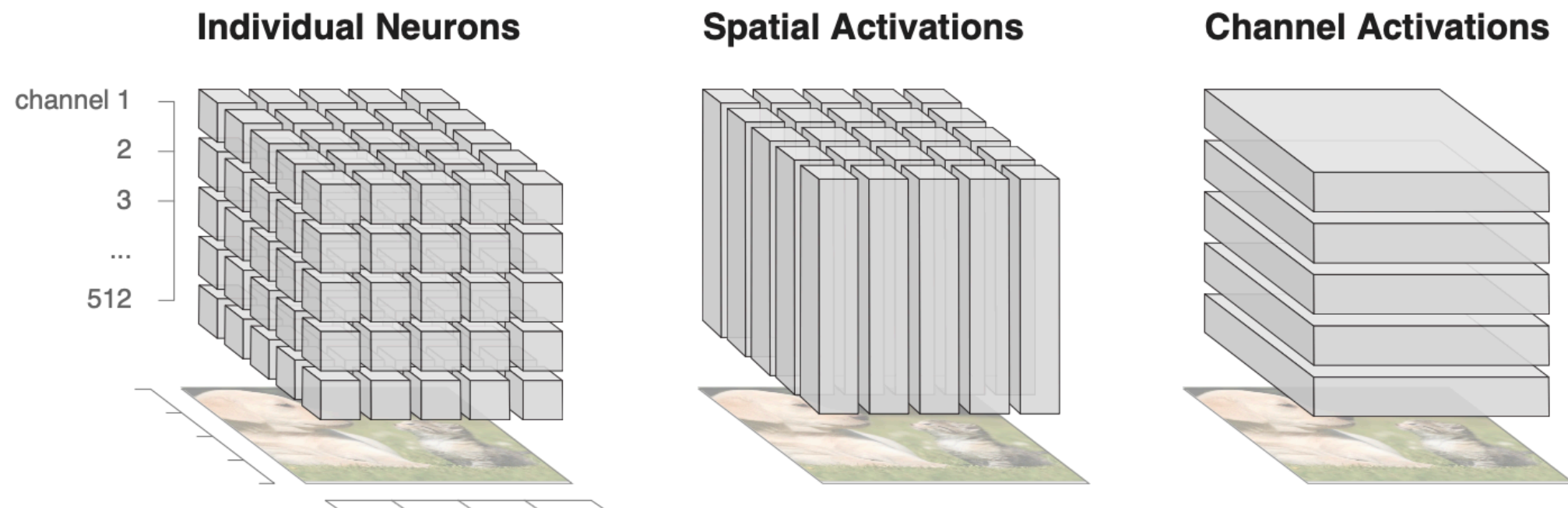
Radar

74%A61M

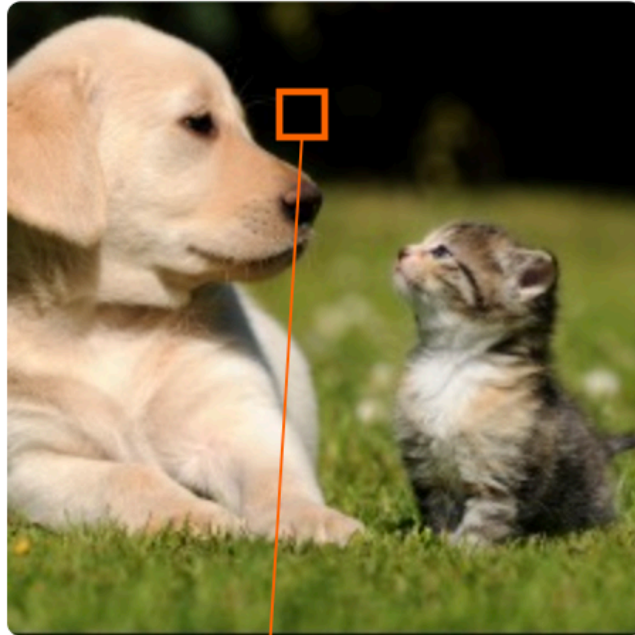
95%B65D

55%A61J

Wyuczone cechy - NN dla obrazów



What Does the Network See?



Semantic dictionaries give us a fine-grained look at an activation: what does each single neuron detect? Building off this representation, we can also consider an activation vector as a whole. Instead of visualizing individual neurons, we can instead visualize the *combination* of neurons that fire at a given spatial location. (Concretely, we optimize the image to maximize the dot product of its activations with the original activation vector.)



Activation Vector

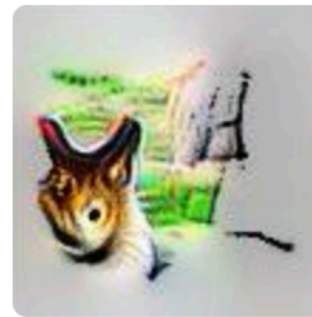
=

432.



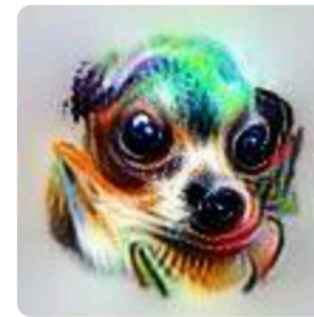
+

273.



+

181.



+

181.



+

...

Channels

INPUT IMAGE

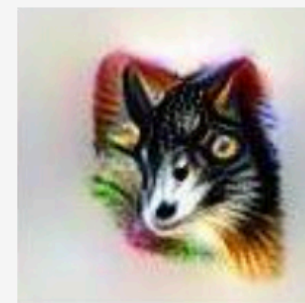
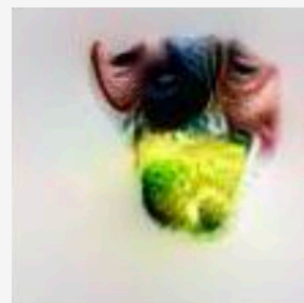
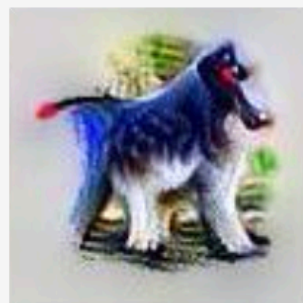
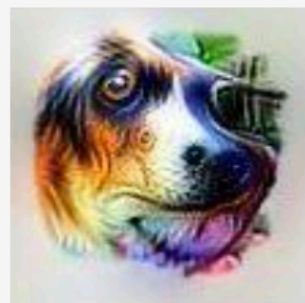
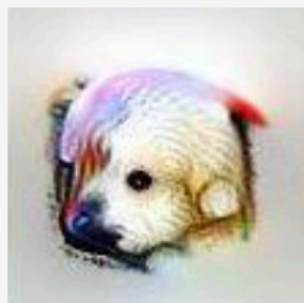
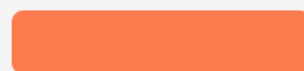
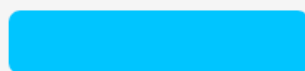
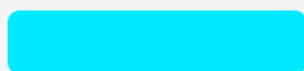
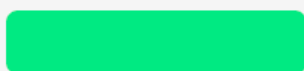


ACTIVATIONS of neuron groups



ed on matrix factorization of mixed4d layer

 6 groups



- <https://christophm.github.io/interpretable-ml-book/>
- <https://arxiv.org/pdf/1606.03490.pdf>
- https://distill.pub/2018/building-blocks/?utm_campaign=Revue%20newsletter&utm_medium=Newsletter&utm_source=Deep%20Learning%20Weekly
- <https://distill.pub/2016/misread-tsne/>
- https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf