

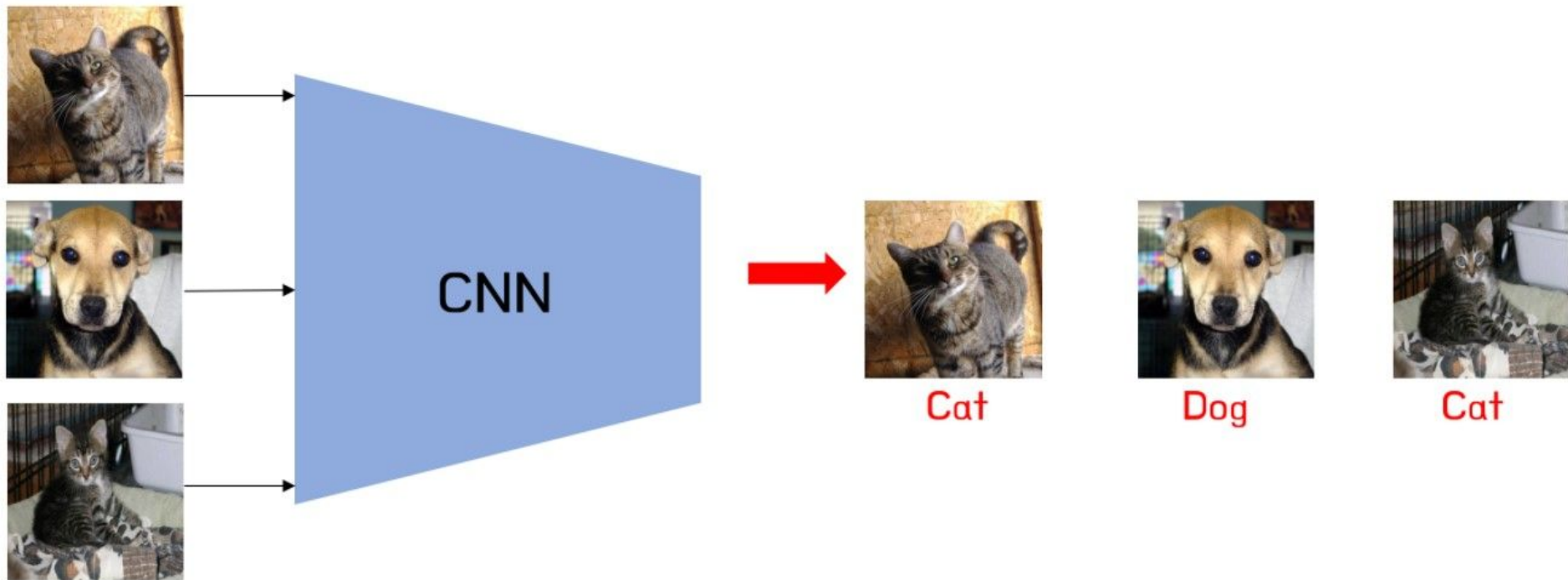
Výuka předmětu Strojové učení

Ing. Jan Mittner, Ph.D.

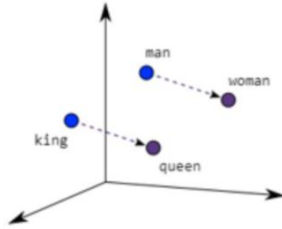
Motivace a cíl

- strojové učení
 - rychle se rozvíjející obor s velkým dopadem
 - zásadní role v rámci většiny systémů umělé inteligence
 - systémy se učí pravidla z dat bez jejich explicitní specifikace programátorem
- na FIS VŠE i VŠPJ spíše teoretické předměty na téma umělé inteligence
- cíl: nový předmět zaměřený na praktické aplikace strojového učení
 - vysvětlit principy strojového učení
 - představit nejpoužívanější algoritmy a modely
 - ukázat typické úlohy řešené pomocí strojového učení
 - naučit se pracovat s nástroji pro praktické aplikace strojového učení
 - vyzkoušet si realizaci projektu strojového učení od začátku do konce

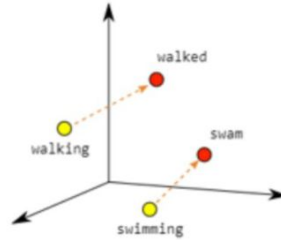
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9



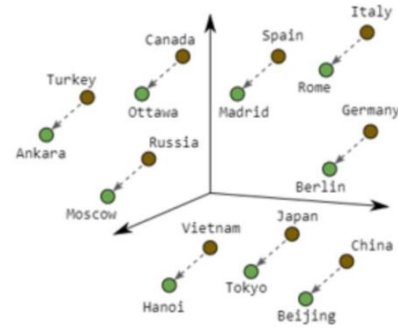
Word2Vec



Male-Female

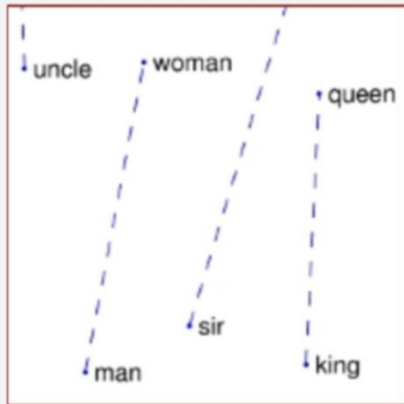


Verb Tense

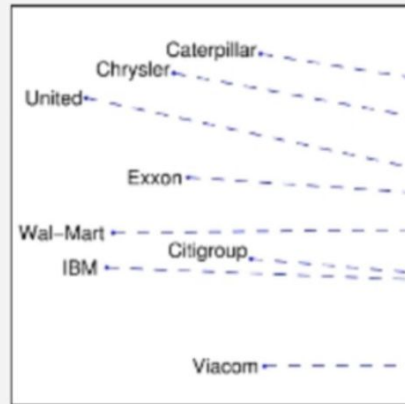


Country-Capital

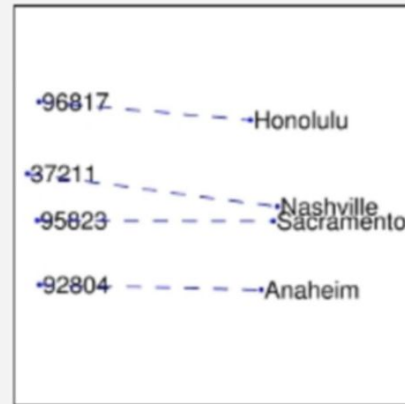
GloVe



man - woman



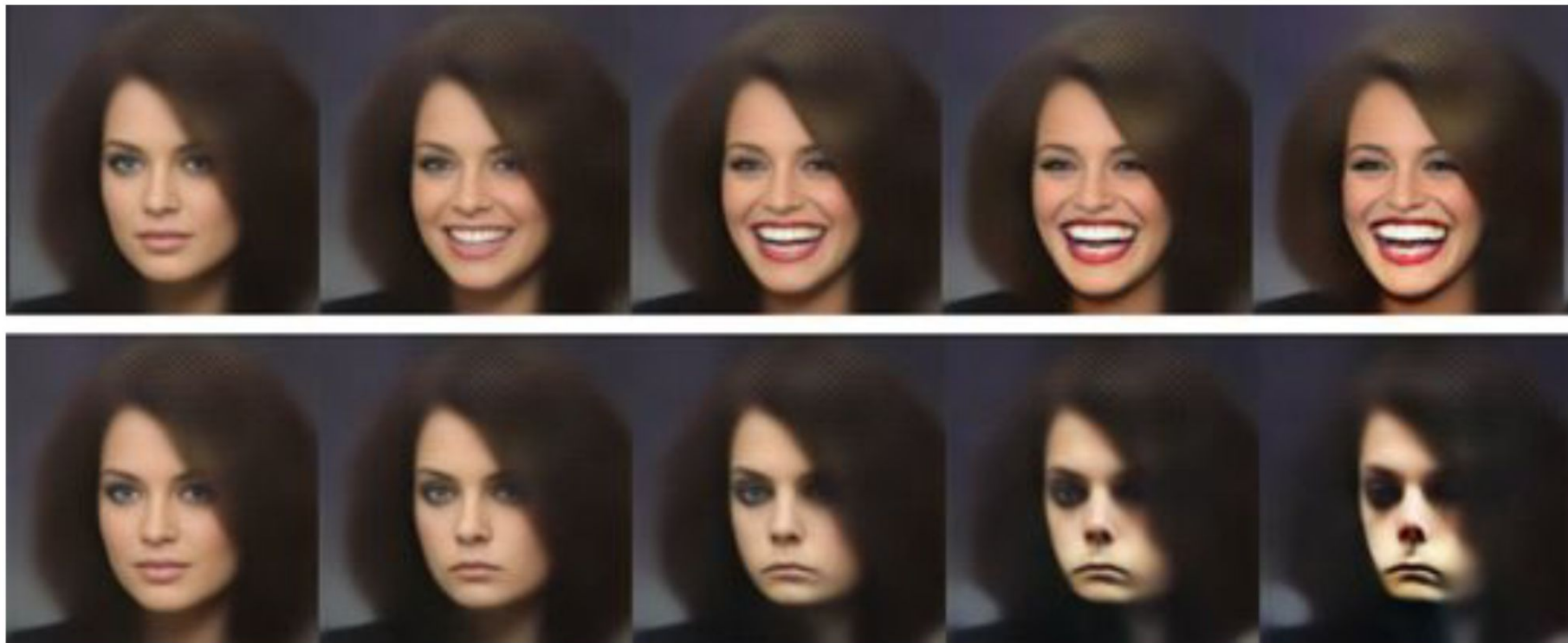
company - ceo



city - zip code



comparative - superlative



<https://arxiv.org/pdf/1609.04468.pdf>

Výzvy při výuce strojového učení

- komplexnost a široká škála disciplín
 - regresní a klasifikační úlohy, neuronové sítě, transformery
 - pochopení problému, získání dat a porozumění jim, předzpracování dat, trénování a ladění modelů, interpretace výsledků, provoz modelů, ...
- široké prerekvizity
 - matematika (zejména lineární algebra)
 - pravděpodobnost a statistika
 - programování
- aktuálnost kurzu
 - dynamický obor s velmi rychlým vývojem
 - za poslední 2-3 roky zásadní posun zejména v oblasti generativní AI

Pedagogické přístupy

- kombinace minimální teorie a maximální praxe
 - srozumitelný teoretický základ vysvětlovaný v rámci možností názorně než skrz vzorečky
 - po teoretickém výkladu rovnou praktické vyzkoušení na konkrétním problému
- použití Jupyter Notebooků pro praktické úlohy
 - každé téma doplněno několika pracovními sešity se vzorově řešenými úlohami, které studenti dále rozvíjejí
- doplňující úlohy formou domácích úkolů za bonusové body
 - (možnost následného výběru snazší semestrální práce)

Jupyter Notebook

- kombinace textu, obrázků, spustitelného kódu a programových výstupů
 - včetně grafických výstupů (grafy, obrázky, ...)
- vynikající pro experimentování, testování, rychlý vývoj, debugging, ...
 - a také pro výuku
- buď jako samostatná aplikace, nebo integrovatelné do Visual Studio Code
- hostovaná varianta Google Colab

Ukázka Jupyter Notebooku

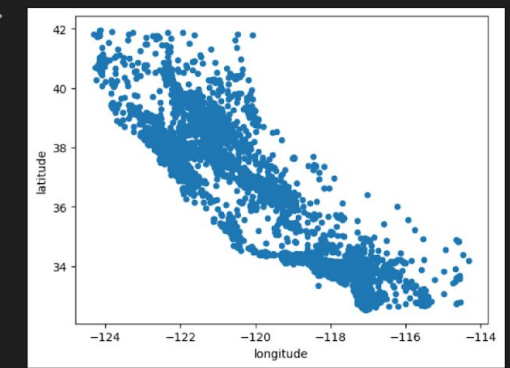
Zkusíme detailně prozkoumat atributy a vztahy mezi nimi

Testovací dataset už máme bokem a budeme podrobněji prozkoumávat trénovací set. Uděláme si kopii, abychom nemodifikovali původní data.

```
housing = train_set.copy()
```

Vizualizace geografických dat

```
housing.plot(kind="scatter", x="longitude", y="latitude")
```



Vytvoření modelu neuronové sítě

```
model = keras.models.Sequential([
    keras.layers.Dense(300, activation="relu", input_shape=[28*28]),
    keras.layers.Dense(100, activation="relu"),
    keras.layers.Dense(10, activation="softmax"),
])
```

```
model.summary()
```

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 300)	235500
dense_1 (Dense)	(None, 100)	30100
dense_2 (Dense)	(None, 10)	1010

Total params: 266,610
Trainable params: 266,610
Non-trainable params: 0

Hustě propojené vrstvy mají hodně parametrů: dimenze vstupu (784) * počet neuronů ve vrstvě (300) + bias (300) + další vrstvy

Hodně parametrů znamená hodně flexibility a také náklonnost k přeučení

```
model.compile(loss="sparse_categorical_crossentropy",
              optimizer="sgd",
              metrics=["accuracy"])
```

Nyní získáme codings pro tyto dva obrázky a pak je budeme interpolovat

```
coding1 = variational_encoder.predict(X_valid[x1:x1+1])
coding2 = variational_encoder.predict(X_valid[x2:x2+2])

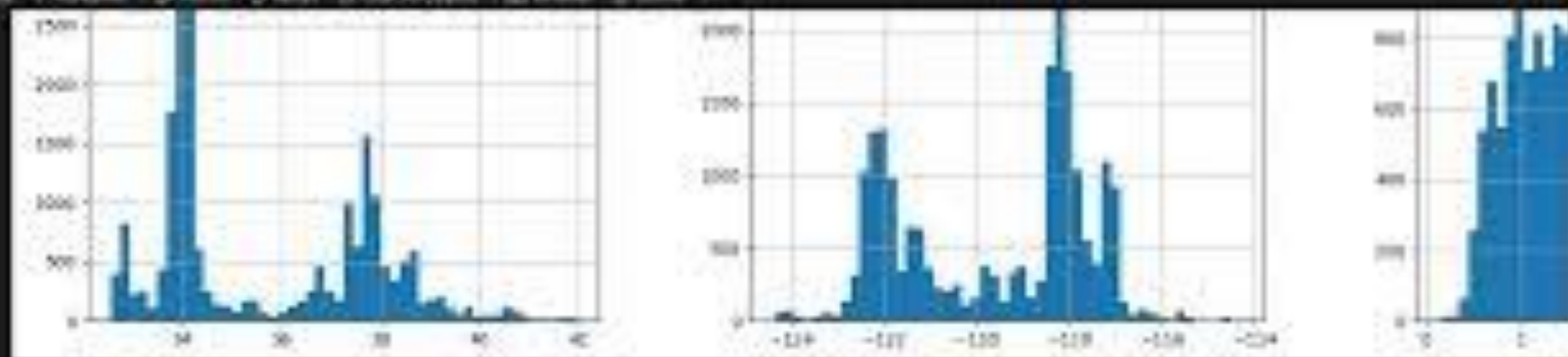
coding1[0], coding2[0]
```

```
interpolated_coding = (coding1[0] + coding2[0]) / 2
interpolated_coding
```

```
new_image = variational_decoder(interpolated_coding.reshape(1,-1)).numpy()
```

```
plot_multiple_images(np.array([X_valid[x1], new_image[0], X_valid[x2]]))
```





Priznanje si dovolite večje pozornosti bolj napreduje, črna vrtilnica pa je

```
housing = pd.DataFrame(columns = ['sqft_living', 'lat', 'lon', 'price'],
                             data = np.random.randn(1000, 4) * 1000,
                             index = np.arange(1, 1001))
```

Jedra so v splošnem bolj povezane, saj so si bolj podobne vrednosti inbo krajši geografski prostori.

Čudilnica povezuje, vrednosti s podobno dolo.

Alternativy k Jupyter Notebooku

Java	JShell	Apache Zeppelin
C++	xeus-cling (Jupyter)	Cling
C#	.NET Interactive	LINQPad

Obsah a struktura kurzu

- **Úvod do strojového učení**
 - hlavní pojmy, historie, současnost, algoritmy, typy úloh
 - prakticky: lineární regrese - predikce HDP podle země.
- **Základní praktiky strojového učení**
 - doporučený pracovní postup při řešení úloh,
 - prakticky: predikce ceny nemovitosti.
- **Klasifikace entit**
 - binární a multiclass klasifikace,
 - prakticky: rozpoznávání psaných číslic.
- **Trénování modelů**
 - metoda Gradient Descent,
 - prakticky: Titanic dataset – klasifikace strukturovaných dat.
- **Pokročilé metody klasifikace**
 - SVM, Rozhodovací stromy, Random Forest,
 - prakticky: SPAM filtr – klasifikace nad nestrukturovanými daty.
- **Učení bez učitele**
 - shlukování, redukce dimenzionality, detekce anomalií,
 - prakticky: systém pro doporučování filmu ke zhlédnutí.
- **Neuronové sítě a Deep Learning**
 - principy a klíčové pojmy, framework Keras,
 - prakticky: binární a multiclass klasifikace, regrese.
- **Deep Learning pro počítačové vidění**
 - konvoluční neuronové sítě,
 - prakticky: rozpoznávání objektů v obraze, rozpoznávání obličejů.
- **Deep Learning pro texty a sekvenční data**
 - word embeddings, předtrénované modely,
 - prakticky: rekurentní NN – predikce teploty z meteo údajů.
- **Generativní Deep Learning**
 - variační autoenkodéry, GAN, transformery,
 - prakticky: generování textu, přenos stylu, generování obrazů.

Zpětná vazba

- předmětové ankety na konci semestru
- FIS VŠE - 20 respondentů
 - obsah a náročnost předmětu: 88/100.
 - překryv s jinými předměty: 97/100.
 - celková spokojenost s předmětem: 98/100.
 - „Co se mi na předmětu líbí“:
 - „Náplň předmětu, domácí úkoly na procvičení a přípravu ke zkoušce, praktické ukázky.“
 - „Jupyter notebooky, praktické ukázky.“
 - „Skvěle odhadnutý poměr mezi teoretickými poznatky a praktickým užitím. Zároveň je skvělé, že jsme si mohli vyzkoušet funkční a praktická řešení, která jsme si během hodiny vyvinuli. Skvělé je i to, že jsme zpracovávali reálné úlohy, ne učebnicové a uměle vykonstruované problémy. Zdůraznění výhod a nevýhod různých přístupů, best practices.“
 - „Přínos do praxe.“
 - „Připravenost přednášek a souborů s kódem, Přístup vyučujícího, volitelné úkoly, praktické využití znalostí“
 - „Co a jak by se na výuce mohlo zlepšit (cvičení i přednášky), abych se více naučil/a“:
 - „Chápu, že se předmět snaží obsáhnout co největší spektrum využití ML, ale přišlo mi, že tempo bylo velmi svižné a od úplné základních věcí jsme se dostali k velmi komplexním dost rychle. V druhé polovině semestru jsem byl často ztracen.“
- VŠPJ - 3 respondenti
 - Ohodnoňte studijní materiály, které jsou vám poskytnuty prostřednictvím e-learningu. Jsou studijní materiály v e-learningu dostačující k tomu, abyste zvládali průběžně samostudium?
 - Odpověď: 3x ano.

Děkuji za pozornost
