Автоэнкодеры

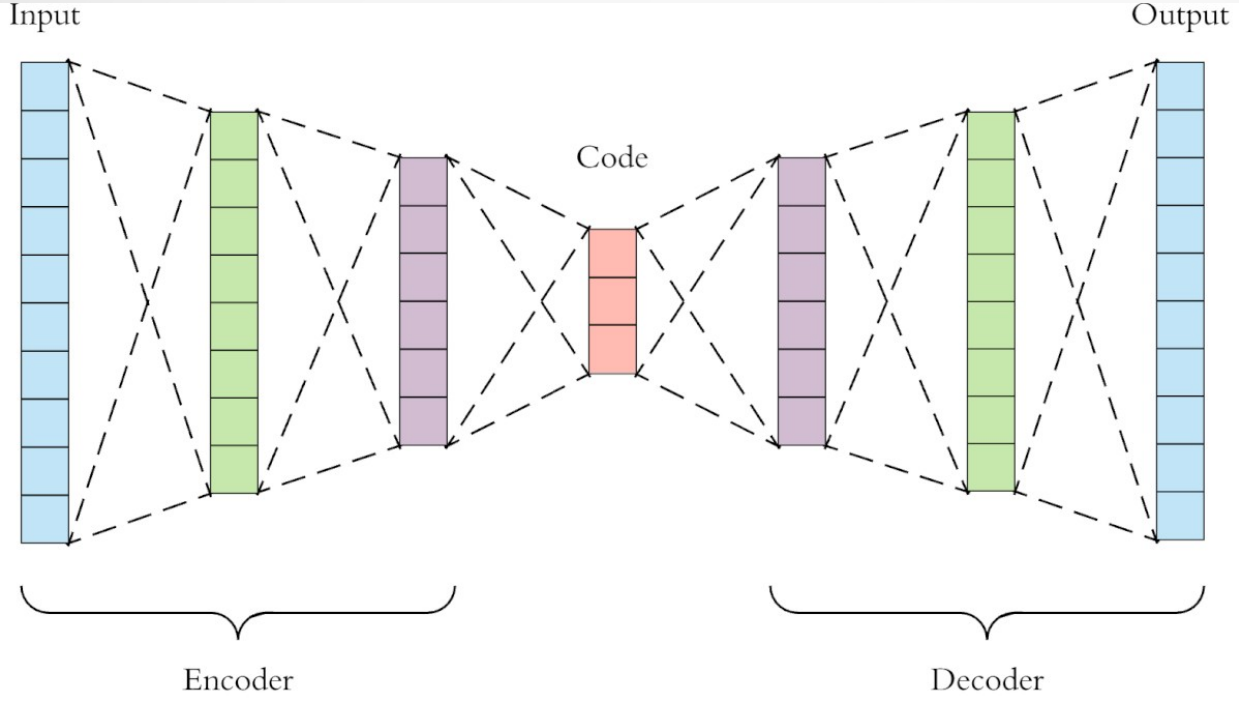
Евгений Борисов

Нейросети

Автоэнкодер

нейронная сеть прямого распространения

вторая половина сети зеркально повторяет первую



при обучении требуем от сети восстановить исходный образ

цель — получить внутренне представление входного образа, (скрытый слой)

https://neurohive.io/ru/osnovy-data-science/avtojenkoder-tipy-arhitektur-i-primenenie/

David Charte et al. A pratial tutorial on autoenoders for nonlinear feature fusion:taxonomy, models, software and guidelines. 2018.

Rumelhart, Hinton, Williams. Learning Internal Representations by Error Propagation.1986.

Нейросети

Задачи для автоэнкодеров

Rumelhart, Hinton, Williams. Learning Internal Representations by Error Propagation.1986.

David Charte et al. A pratial tutorial on autoenoders for nonlinear feature fusion:taxonomy, models, software and guidelines. 2018.

- Генерация признаков (feature generation)
- Снижение размерности (dimensionality reduction)
- Сжатие данных с минимальными потерями точности
- Более эффективное решение задач обучения с учителем в новом признаковом пространстве
- Обучаемая векторизация объектов, встраиваемая в более глубокие нейросетевые архитектуры
- Послойное предобучение многослойных сетей
- Генерация синтетических объектов, похожих на реальные

Нейросети

Постановка задачи автоэнкодера

$X^\ell = \{x_1, \ldots, x_\ell\}$ — обучающая выборка

$f: X \to Z$ — кодировщик (encoder), кодовый вектор $z = f(x, \alpha)$

$g: Z \to X$ — декодировщик (decoder), реконструкция $\hat{x} = g(z, \beta)$

Суперпозиция $\hat{x} = g(f(x))$ должна восстанавливать исходные x_i:

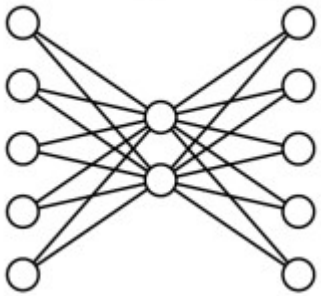
$$\mathscr{L}_{\text{AE}}(\alpha, \beta) = \sum_{i=1}^{\ell} \mathscr{L}\big(g(f(x_i, \alpha), \beta), x_i\big) \to \min_{\alpha,\beta}$$

Квадратичная функция потерь: $\mathscr{L}(\hat{x}, x) = \|\hat{x} - x\|^2$

изменение размерности

извлечение признаков

снижение размерности



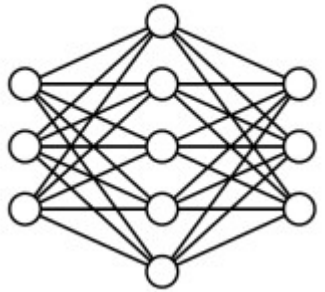
Пример 1. Линейный автокодировщик: $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$

$$f(x, A) = \underset{m \times n}{A}\, x, \qquad g(z, B) = \underset{n \times m}{B}\, z$$

Пример 2. Двухслойная сеть с функциями активации σ_f, σ_g:

$$f(x, A) = \sigma_f(Ax + a), \qquad g(z, B) = \sigma_g(Bz + b)$$

повышение размерности



Нейросети

Шумоподавляющий автоэнкодер (Denoising AE)

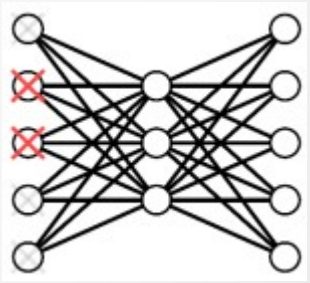
P. Vinent, H. Larohelle, Y. Bengio, P.-A. Manzagol. Extrating and composing robust features with denoising autoenoders. ICML-2008.

при обучении на входной образ накладываем шум

требуем от сети восстановить исходный образ

Устойчивость кодовых векторов z_i относительно шума в x_i:

$$\mathscr{L}_{\text{DAE}}(\alpha,\beta) = \sum_{i=1}^{\ell} \mathsf{E}_{\tilde{x}\sim q(\tilde{x}|x_i)} \mathscr{L}\big(g(f(\tilde{x},\alpha),\beta),x_i\big) \to \min_{\alpha,\beta}$$



Вместо вычисления $\mathsf{E}_{\tilde{x}}$ в методе SGD объекты x_i сэмплируются и зашумляются по одному: $\tilde{x} \sim q(\tilde{x}|x_i)$. Варианты зашумления:

- $\tilde{x} \sim \mathcal{N}(x_i, \sigma^2 I)$ — гауссовский шум
- обнуление компонент вектора x_i с вероятностью p_0:

Нейросети

Автоэнкодер для обучения на размеченных данных

Dor Bank, Noam Koenigstein, Raja Giryes. Autoenoders. 202

Данные: неразмеченные $(x_i)_{i=1}^{\ell}$, размеченные $(x_i, y_i)_{i=\ell+1}^{\ell+k}$

Совместное обучение кодировщика, декодировщика и предсказательной модели (классификации, регрессии или др.):

$$\sum_{i=1}^{\ell} \mathscr{L}\big(g(f(x_i,\alpha),\beta),x_i\big) + \lambda \sum_{i=\ell+1}^{\ell+k} \tilde{\mathscr{L}}\big(\hat{y}(f(x_i,\alpha),\gamma),y_i\big) \to \min_{\alpha,\beta,\gamma}$$

$z_i = f(x_i,\alpha)$ — кодировщик

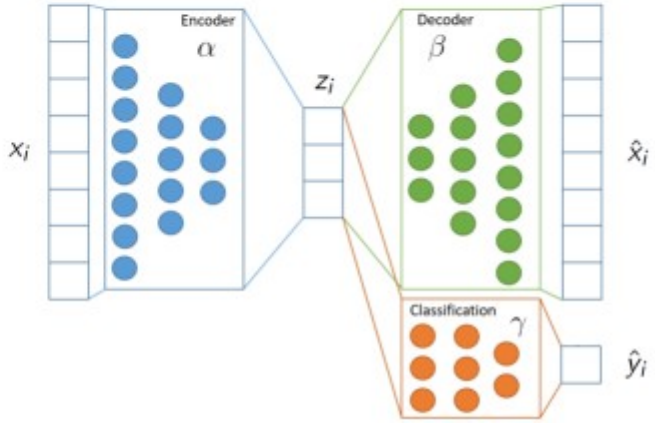
$\hat{x}_i = g(z_i,\beta)$ — декодировщик

$\hat{y}_i = \hat{y}(z_i,\gamma)$ — предиктор

Функции потерь:

$\mathscr{L}(\hat{x}_i, x_i)$ — реконструкция

$\tilde{\mathscr{L}}(\hat{y}_i, y_i)$ — предсказание



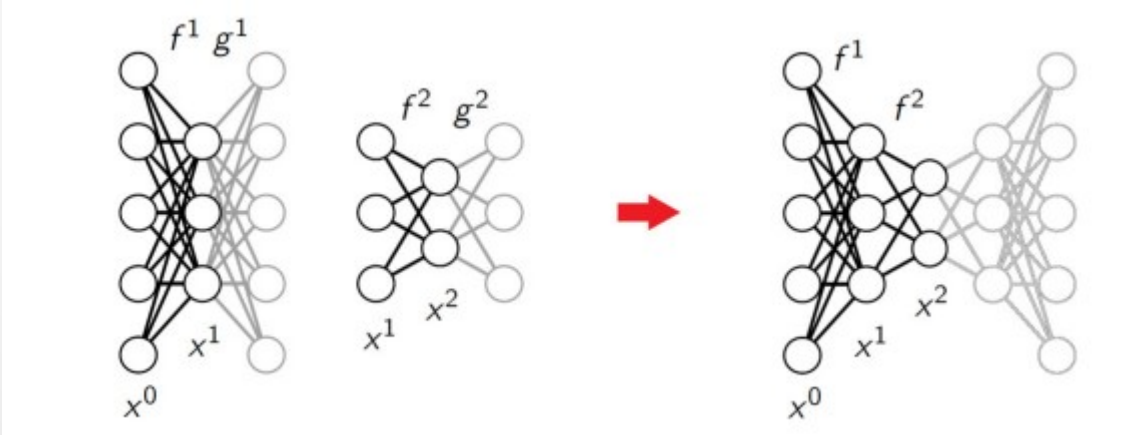
Нейросети

Многослойный автоэнкодер (Staked AE)

Y. Bengio et al. Greedy layer-wise training of deep networks. NIPS 2007.

Послойное обучение: $x^h = f^h(x^{h-1}, \alpha^h)$, $x \equiv x^0$, $z \equiv x^H$

- каждая пара f^h, g^h обучается по выборке $\{x_1^{h-1}, \ldots, x_\ell^{h-1}\}$
- декодировщик g^h отбрасывается
- однослойные $f^1, \ldots, f^H$ соединяются в H-слойный



Тонкая настройка (fine tuning): результат послойного обучения используется как начальное приближение для BackProp

Нейросети

Пример

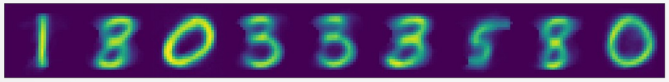
снижение размерности, извлечение признаков

размер входного (и выходного слоя) - 784 = 28x28
размер представления (скрытого) слоя — 2

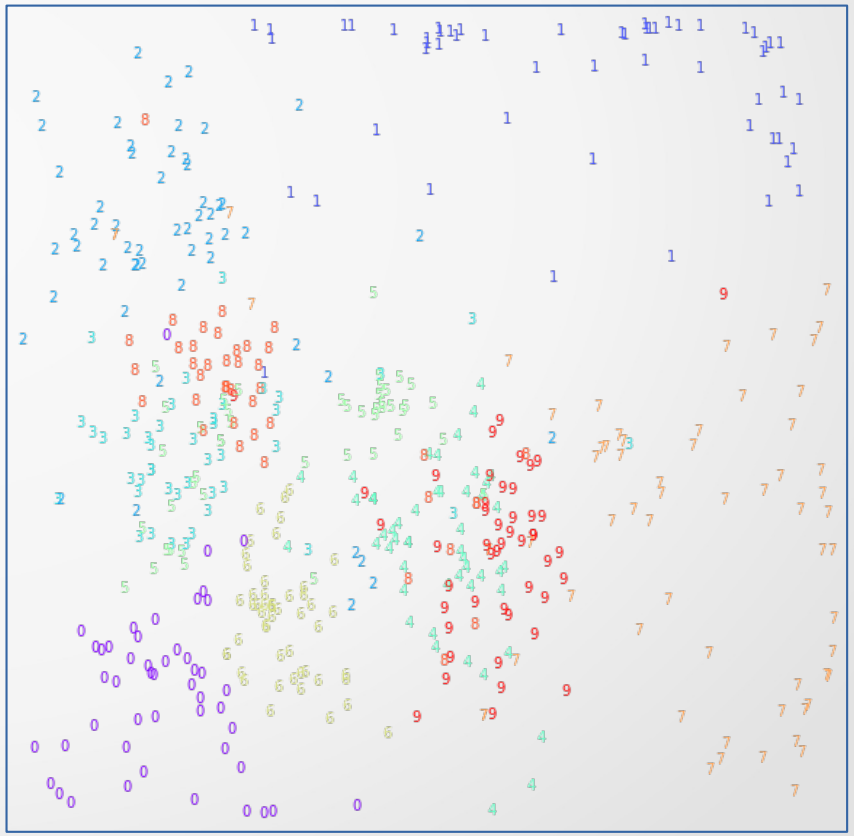
датасет MNIST

оригинал

восстановленный



карта расположения объектов в 2D пространстве признаков



Нейросети

Интерпретация

Изображения цифр mnist можно рассматривать как элементы 28*28=784-мерного пространства.

Среди всех изображений 28x28, изображения цифр занимают небольшую часть, остальное это шум.

Для одной выбранной цифры в 784-мерном пространстве можно найти кривую, все точки некоторой области вдоль этой кривой это цифры.

Т.е. в пространстве всех изображений есть подпространство меньшей размерности с цифрами, которое и находит атоэнкодер.

Нейросети: литература

git clone https://github.com/mechanoid5/ml_lectorium.git

Воронцов К. В.
Прикладные модели машинного обучения. 2021.
Лекция 2: Обучение без учителя.
https://www.youtube.com/watch?v=wfbe2yaXAkI

Rumelhart, Hinton, Williams. Learning Internal Representations by Error Propagation.1986.

David Charte et al. A pratial tutorial on autoenoders for nonlinear feature fusion:taxonomy, models, software and guidelines. 2018.

Михаил Сурцуков
Manifold learning и скрытые (latent) переменные
https://habr.com/ru/post/331500/