

Few-Shot Learning

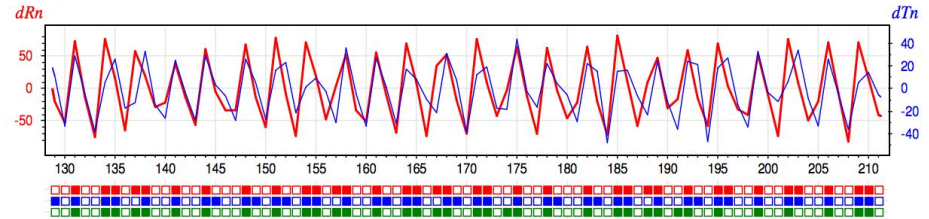
Козлов Василий

Мотивация

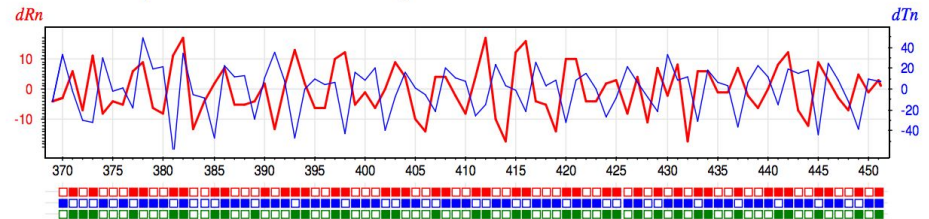
- дорогая разметка данных
- отсутствие данных
- изменение данных
- человекоподобное обучение

Приращения dR_n , dT_n , $d\alpha_n$ в последовательных кардиоциклах n

Здоровый:



Больной (язвенная болезнь):



Определение

FSL - область машинного обучения, отличительной чертой которой, является наличие небольшого числа доступных для обучения данных.

Основные сценарии использования FSL:

- уменьшение ресурсозатратности сбора данных и вычислительной сложности
- обучение для редких случаев
- выступление в качестве первопроходца для человекоподобного обучения

Постановка проблемы

Эмпирический риск $R(h) = \int \ell(h(x), y) dp(x, y) = \mathbb{E}[\ell(h(x), y)].$

Оценка эмпирического риска $R_I(h) = \frac{1}{n} \sum_{i=1}^I \ell(h(x^{(i)}), y^{(i)})$

Разложим матожидание разницы полученного эмпирического риска и точки минимума.
Пусть:

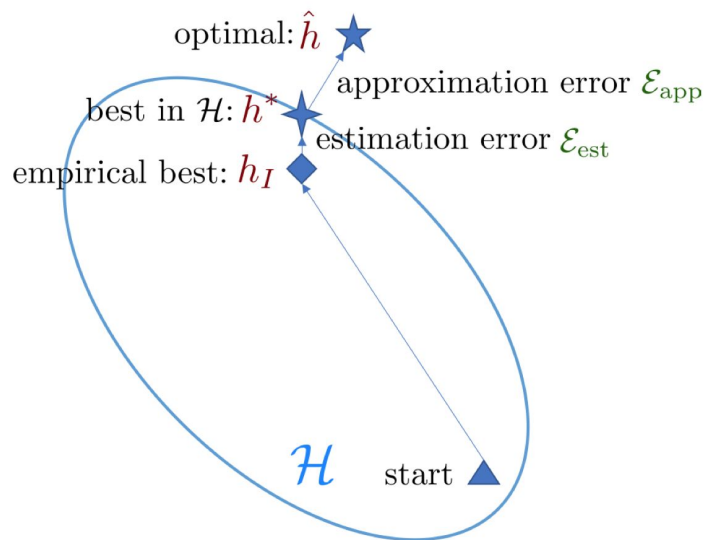
- $\hat{h} = \arg \min_f R(h)$, where R attains its minima;
- $h^* = \arg \min_{h \in \mathcal{H}} R(h)$, where R is minimized with respect to $h \in \mathcal{H}$;
- $h_I = \arg \min_{h \in \mathcal{H}} R_I(h)$, where R_I is minimized with respect to $h \in \mathcal{H}$.

$$\mathbb{E}[R(h_I) - R(\hat{h})] = \underbrace{\mathbb{E}[R(h^*) - R(\hat{h})]}_{\mathcal{E}_{\text{app}}(\mathcal{H})} + \underbrace{\mathbb{E}[R(h_I) - R(h^*)]}_{\mathcal{E}_{\text{est}}(\mathcal{H}, I)}$$

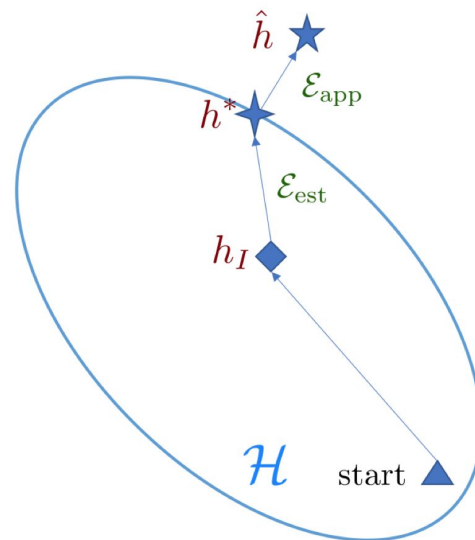
Постановка проблемы

В идеале: $\mathcal{E}_{\text{est}}(\mathcal{H}, \infty) = \lim_{I \rightarrow \infty} \mathbb{E}[R(h_I) - R(h^*)] = 0$

Основная проблема: плохо оцениваем эмпирический риск

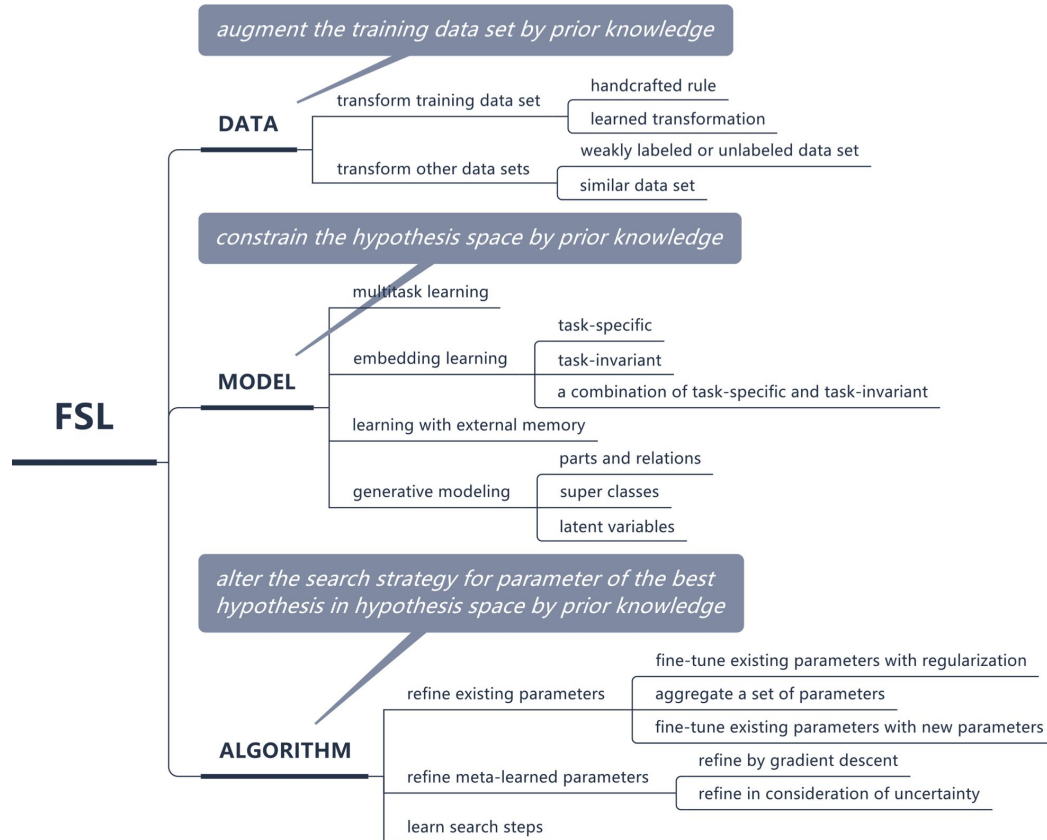


а) Много данных



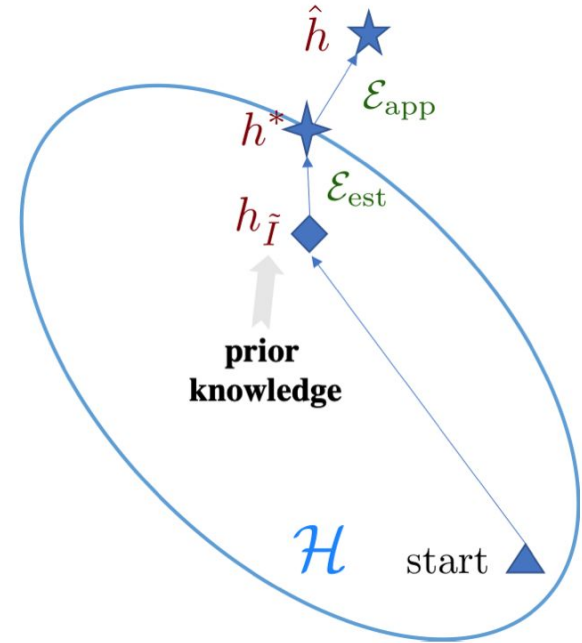
б) Мало данных

Пути решения



Data

Используя априорные знания мы преобразуем D_{train}

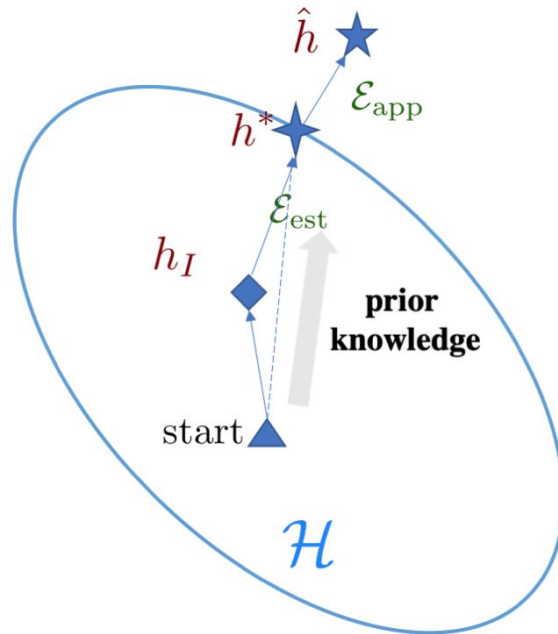


prior knowledge	transformation		
	input	transformer	output
handcrafted rule	original (x, y)	handcrafted rule on x	(transformed x, y)
learned transformation	original (x, y)	learned transformation on x	(transformed x, y)
weakly labeled or unlabeled data set	weakly labeled or unlabeled x	predictor h trained by D^{train}	$(x, \text{output predicted by } h)$
similar data set	sample from similar data set	aggregate new x and y by weighted average of samples of similar data set	aggregated sample

Algorithm

Предварительное знание изменяет стратегию поиска предоставляя:

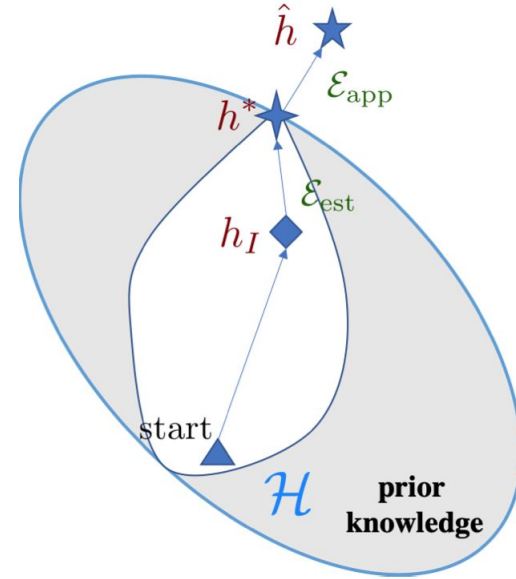
- хорошую начальную точку
- непосредственно шаги поиска.



strategy	prior knowledge	how to search θ of the h^* in \mathcal{H}
refine existing parameters θ^0	learned θ^0	refine θ^0 by D^{train}
refine meta-learned θ	meta-learner	refine θ by D^{train}
learn search steps	meta-learner	use search steps provided by the meta-learner

Model

Сужая область гипотез мы повышаем качество оценки эмпирического риска



strategy	prior knowledge	how to constrain \mathcal{H}
multitask learning	other T 's with their data sets D 's	share parameter
embedding learning	embedding learned from/together with other T 's	project samples to a smaller embedding space where similar and dissimilar samples can be easily discriminated
learning with external memory	embedding learned from other T 's to interact with memory	refine samples by D^{train} stored in memory
generative modeling	prior model learned from other T 's	restrict the form of distribution

Multi-task learning

Определение.

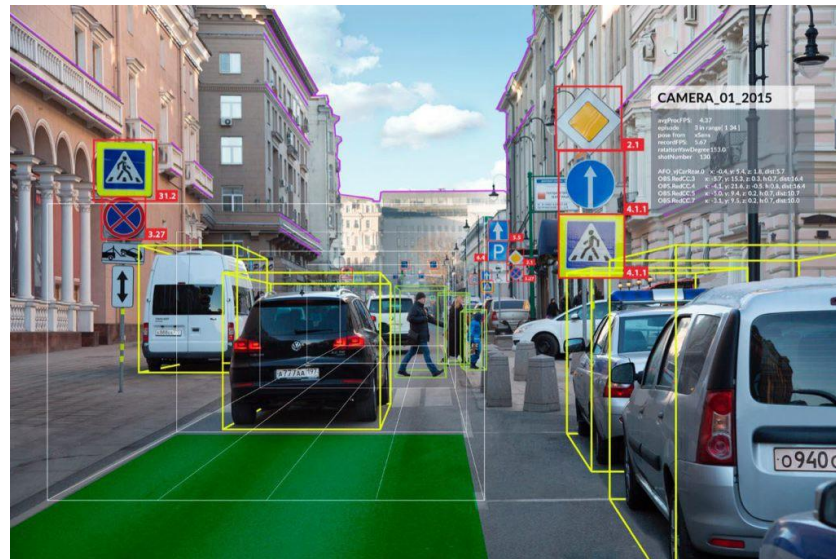
Multi-task learning - область машинного обучения, в которой несколько задач решаются одновременно. Использование сходств и различий помогает улучшить качество предсказаний подмножества моделей.

Применимость к FSL.

- target task
- source tasks

Мотивация.

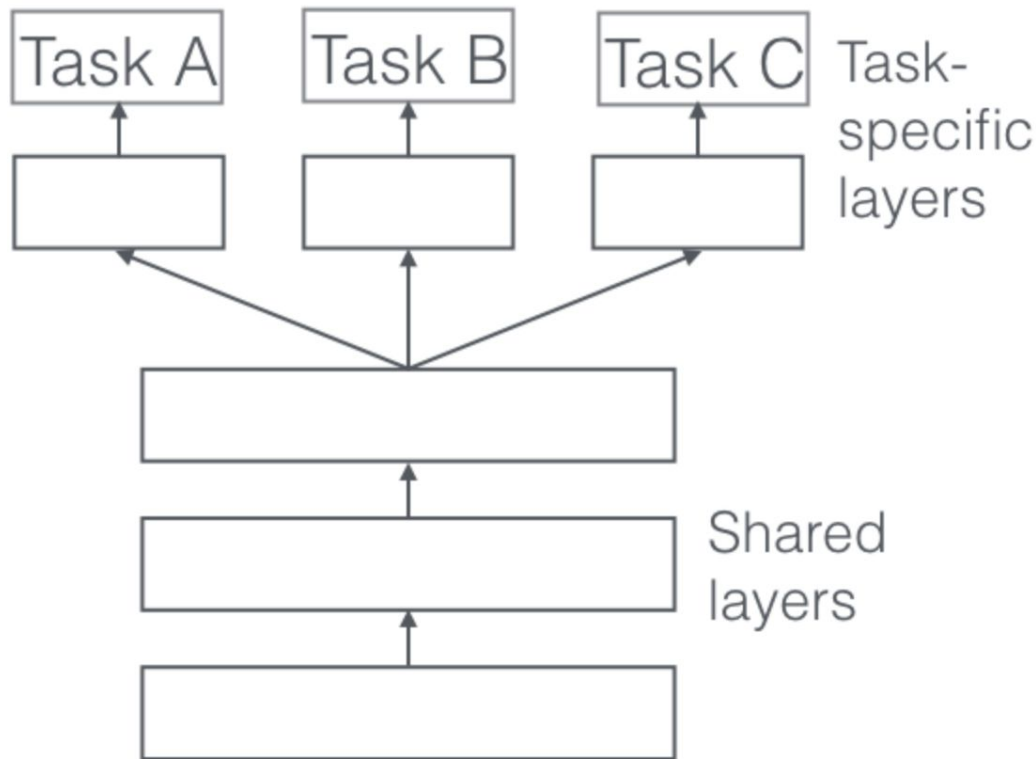
- человекоподобное обучение
- сходство с transfer learning



Hard parameter sharing

Структура

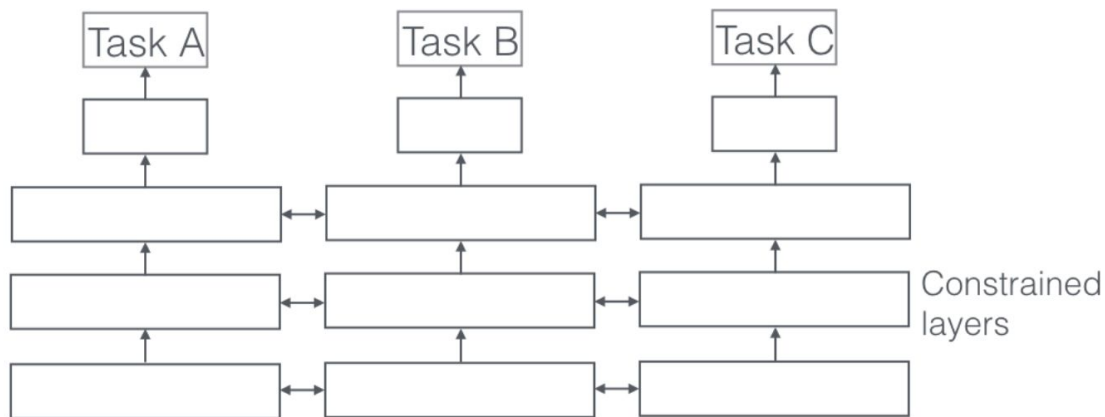
- несколько общих слоёв
- несколько личных слоёв для каждой задачи
- применительно к FSL считаем что target task может обновлять только свои слои, тогда как остальные задачи могут обновлять все слои



Soft parameter sharing

Структура

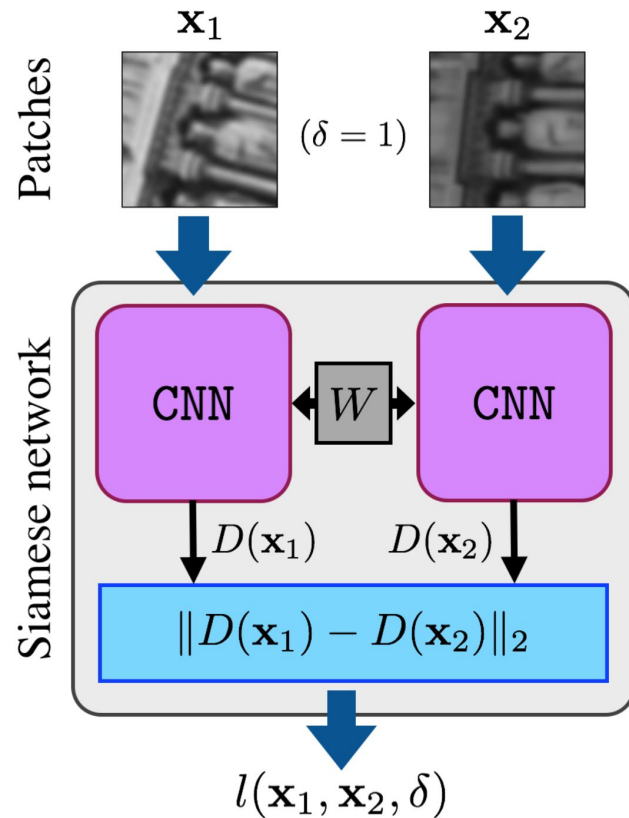
- у каждой задачи своя модель
- используем регуляцию для схожести весов



Сиамские сети

Структура

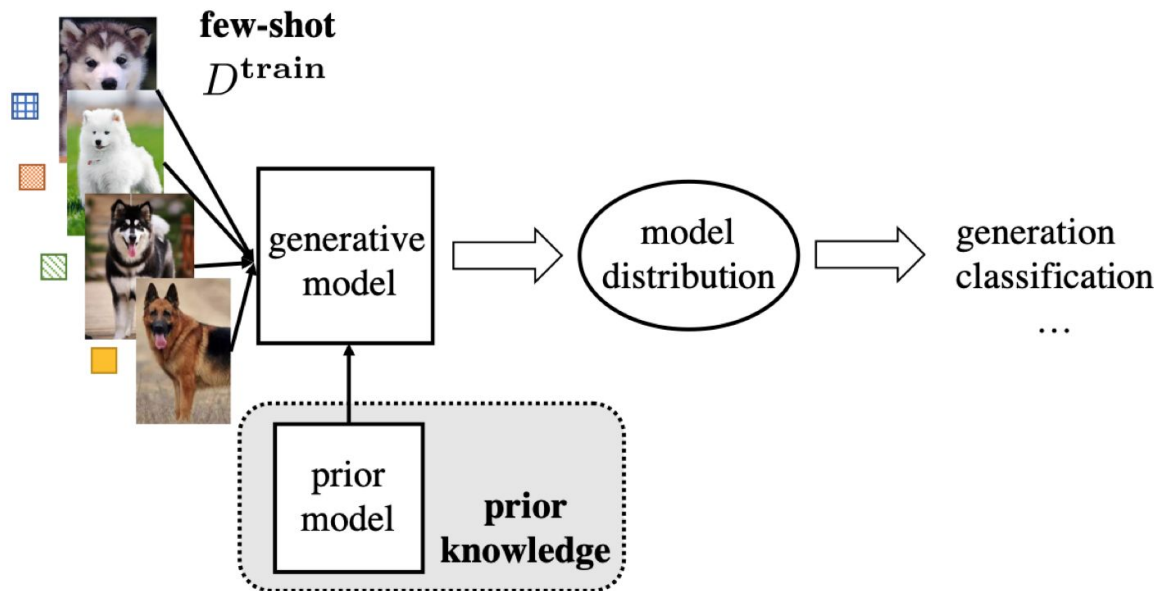
- две CNN
- веса одинаковы
- связываются в конце функцией энергии
- логистическая функция потерь



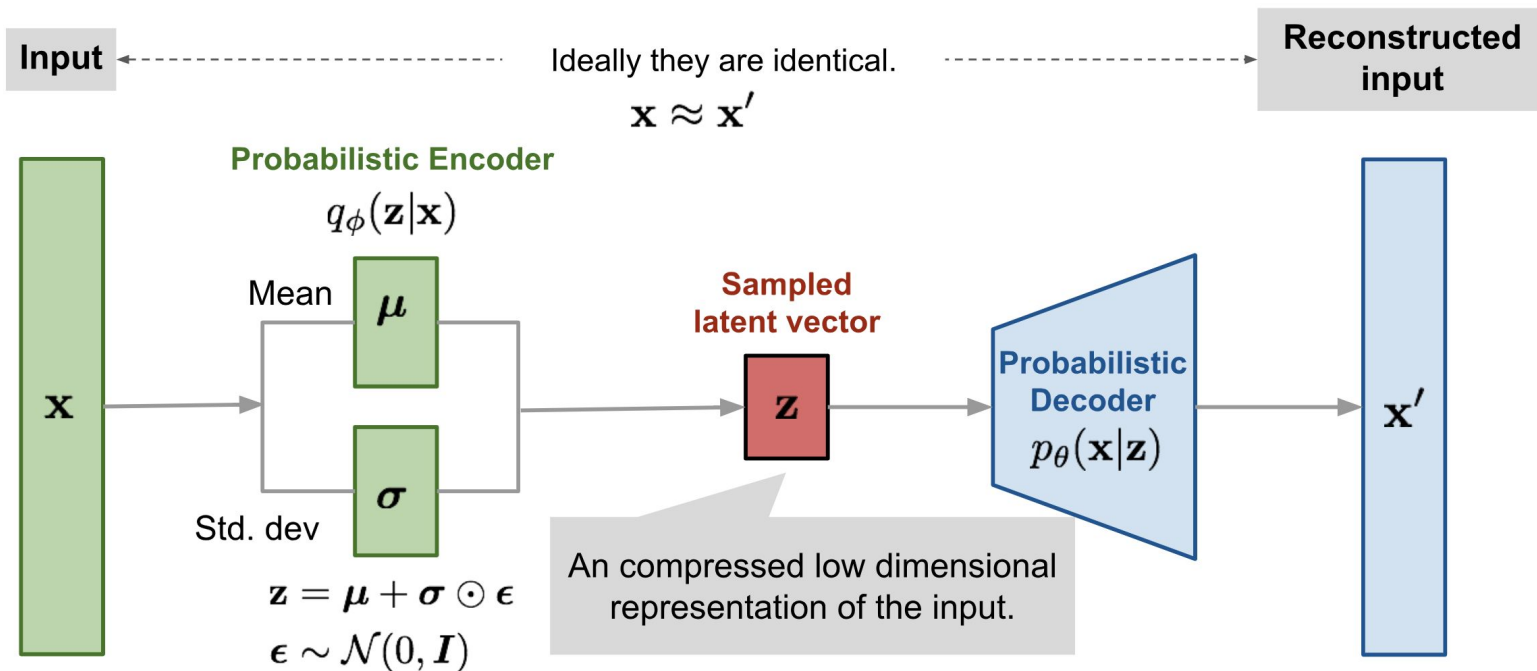
Генеративные модели

Стратегия

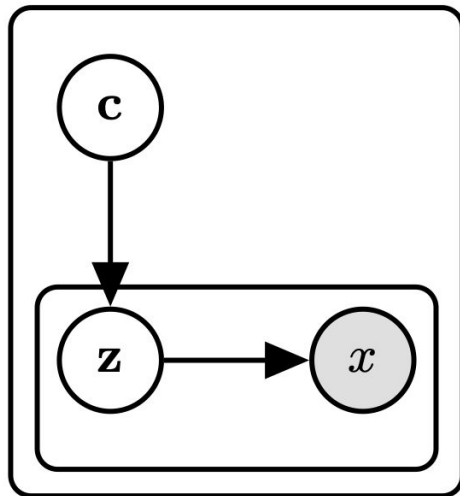
- используя априорные знания и наш дата сет хотим предсказывать распределение
- будем использовать априорные модели натренированные на других дата-сетах
- генеративные методы обновляют распределение для нашего дата сета



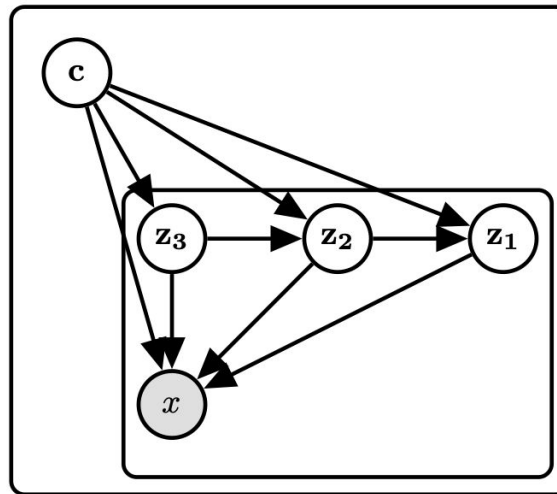
Variation autoencoder



Neural statistician



а) базовая модель



б) neural statistician

Algorithm 4 K -way few-shot classification

$D_0, \dots, D_K \leftarrow$ sets of labelled examples for each class

$x \leftarrow$ datapoint to be classified

$N_x \leftarrow q(c|x; \phi)$ {approximate posterior over c given query point}

for $i = 1$ **to** K **do**

$N_i \leftarrow q(c|D_i; \phi)$

end for

$\hat{y} \leftarrow \operatorname{argmin}_i D_{KL}(N_i \| N_x)$

Источники

<https://arxiv.org/pdf/1904.05046.pdf> FSL и связь всех перечисленных методов с ним

<https://arxiv.org/pdf/1707.08114.pdf> MTL

<https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf> Сиамские сети

<https://arxiv.org/pdf/1606.02185.pdf> Neural statistician

Вопросы

- 1) Опишите основные сценарии использования FSL
- 2) Опишите структуру MTL для случаев hard и soft parameter sharing
- 3) Опишите стратегию генеративных методов в FSL