

Few-Shot Learning

Малафеев Михаил

Few-Shot Learning. Что это?

FSL – область машинного обучения, особенная тем, что имеется малое количество данных для обучения.

- Малое количество размеченных данных
- Обучение происходит аналогично тому, как учится человек
- Возможно, имеем какие-то знания о похожих данных

Few-Shot Learning. Примеры

- ✓ Меньше требований к сбору данных и их разметке
- ✓ Подходит для редких случаев
- ✓ Обучается как человек

task T	experience E		performance P
	supervised information	prior knowledge	
character generation [76]	a few examples of new character	pre-learned knowledge of parts and relations	pass rate of visual Turing test
drug toxicity discovery [4]	new molecule's limited assay	similar molecules' assays	classification accuracy
image classification [70]	a few labeled images for each class of the target T	raw images of other classes, or pre-trained models	classification accuracy

Few-Shot Learning. Функция ошибки

- $p(x, y)$ — реальное распределение выхода y для входа x
- $l(\hat{y}, y)$ — функция потерь
- $h(x)$ — предположение

Ожидаемый риск $R(h) = \int l(h(x), y) dp(x, y) = E[l(h(x), y)]$

Эмпирический риск $R_I(h) = \frac{1}{I} \sum_{i=1}^I l(h(x_i), y_i)$

Few-Shot Learning. Разложение ошибки

- H — пространство гипотез
- \hat{h} — гипотеза, при которой минимизируем R
- h^* — гипотеза из H , при которой минимизируем R
- h_I — гипотеза из H , при которой минимизируем R_I

$$E[R(h_I) - R(\hat{h})] = E[R(h^*) - R(\hat{h})] - E[R(h_I) - R(h^*)]$$

Few-Shot Learning. Разложение ошибки

$$\mathbb{E}[R(h_I) - R(\hat{h})] = \mathbb{E}[R(h^*) - R(\hat{h})] - \mathbb{E}[R(h_I) - R(h^*)]$$

- $\mathbb{E}[R(h^*) - R(\hat{h})]$ — ошибка приближения (approximation error)
- $\mathbb{E}[R(h_I) - R(h^*)]$ — ошибка обучения (estimation error)

Few-Shot Learning. Задача

Основная задача: уменьшить эмпирический риск

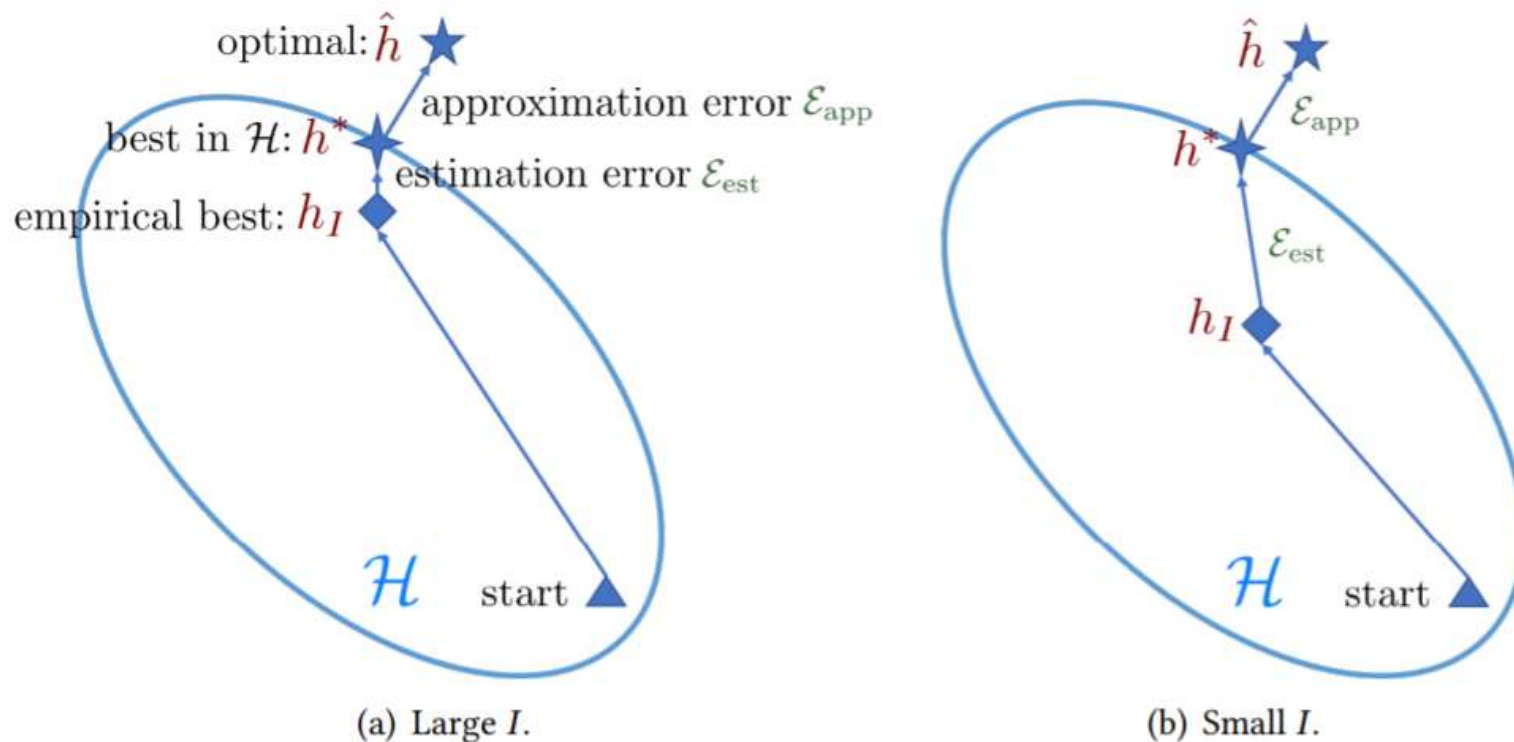
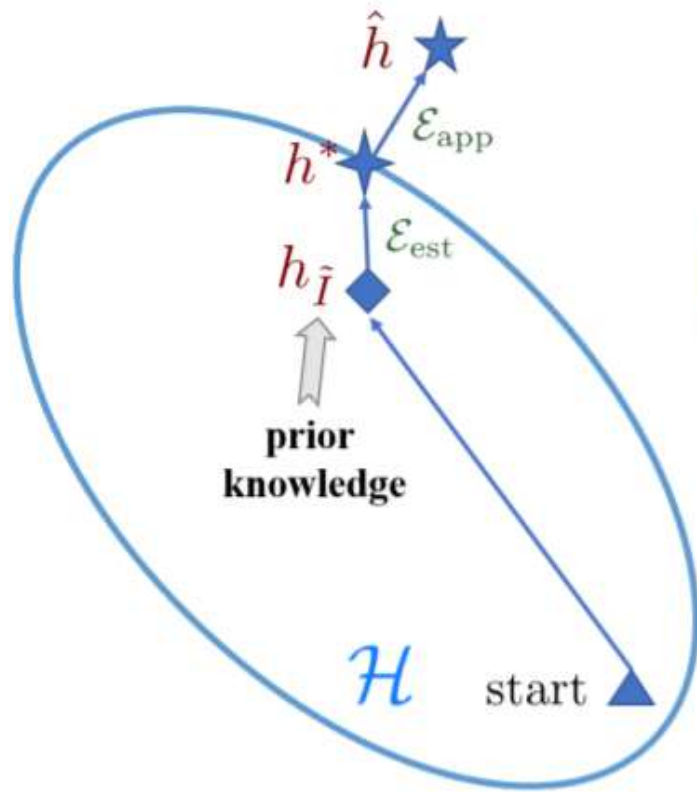
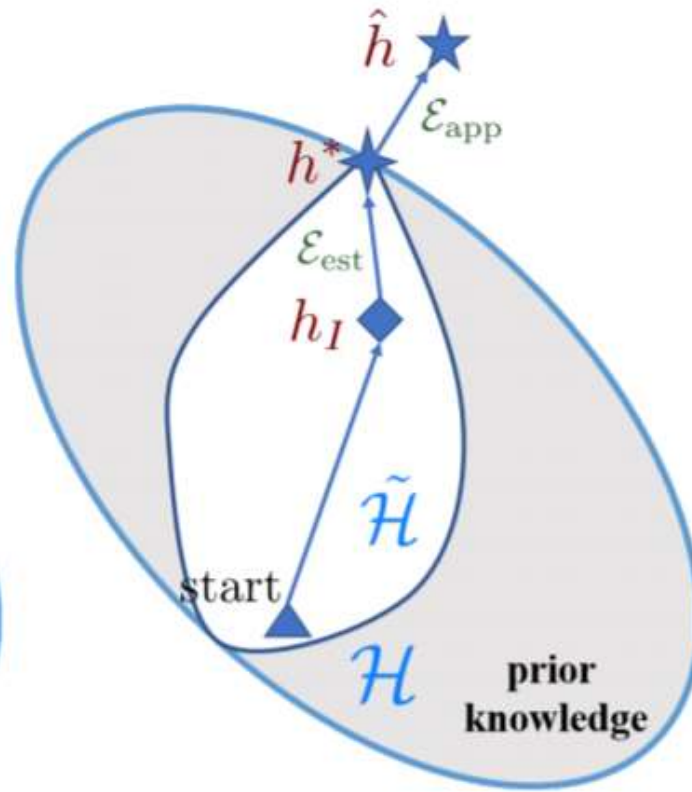


Fig. 1. Comparison of learning with sufficient and few training samples.

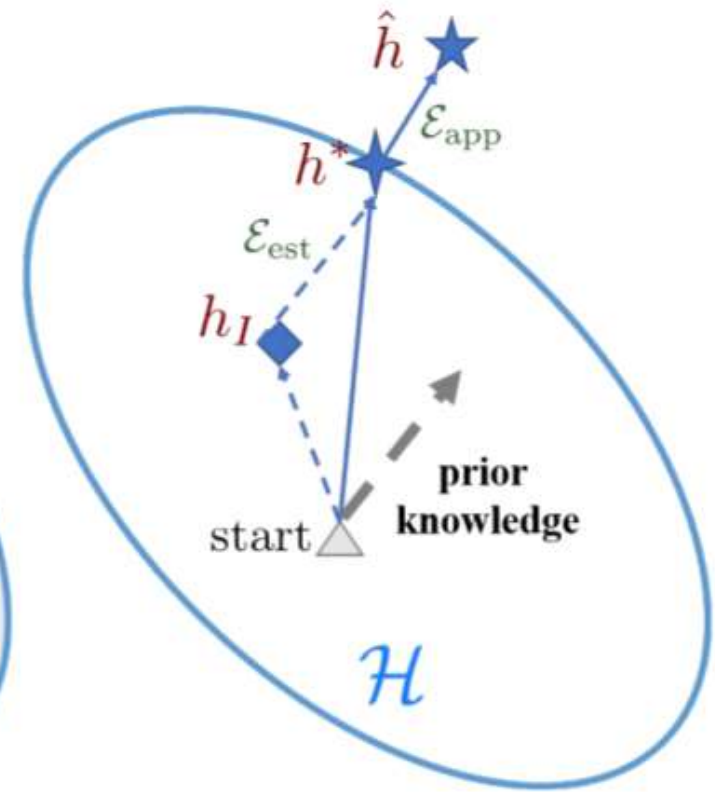
Few-Shot Learning. Уменьшение ошибки



(a) Data.



(b) Model.



(c) Algorithm.

Уменьшение ошибки. Data

- Аугментация данных (поворот, сдвиг, приближение...)
- Преобразование данных с использованием априорных знаний

category	input (x, y)	transformer t	output (\tilde{x}, \tilde{y})
transforming samples from D_{train}	original (x_i, y_i)	learned transformation function on x_i	$(t(x_i), y_i)$
transforming samples from a weakly labeled or unlabeled data set	weakly labeled or unlabeled $(\bar{x}, -)$	a predictor trained from D_{train}	$(\bar{x}, t(\bar{x}))$
transforming samples from similar data sets	samples $\{(\hat{x}_j, \hat{y}_j)\}$ from similar data sets	an aggregator to combine $\{(\hat{x}_j, \hat{y}_j)\}$	$(t(\{\hat{x}_j\}), t(\{\hat{y}_j\}))$

Уменьшение ошибки. Algorithms

- Выбор лучшего начального приближения
- Изменение шагов поиска оптимального решения

strategy	prior knowledge	how to search θ of the h^* in \mathcal{H}
refining existing parameters	learned θ_0	refine θ_0 by D_{train}
refining meta-learned parameters	meta-learner	refine θ_0 by D_{train}
learning the optimizer	meta-learner	use search steps provided by the meta-learner

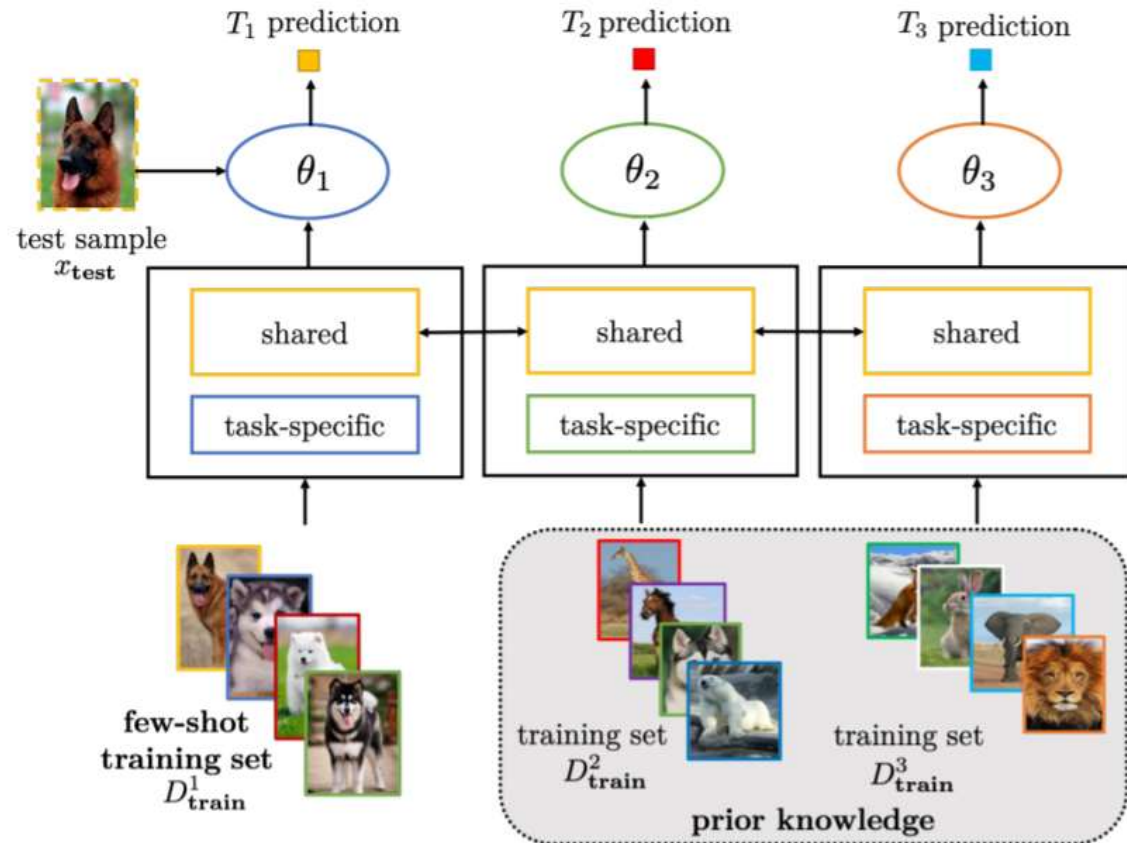
Уменьшение ошибки. Models

С помощью априорных знаний ограничиваем область гипотез \mathcal{H}

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

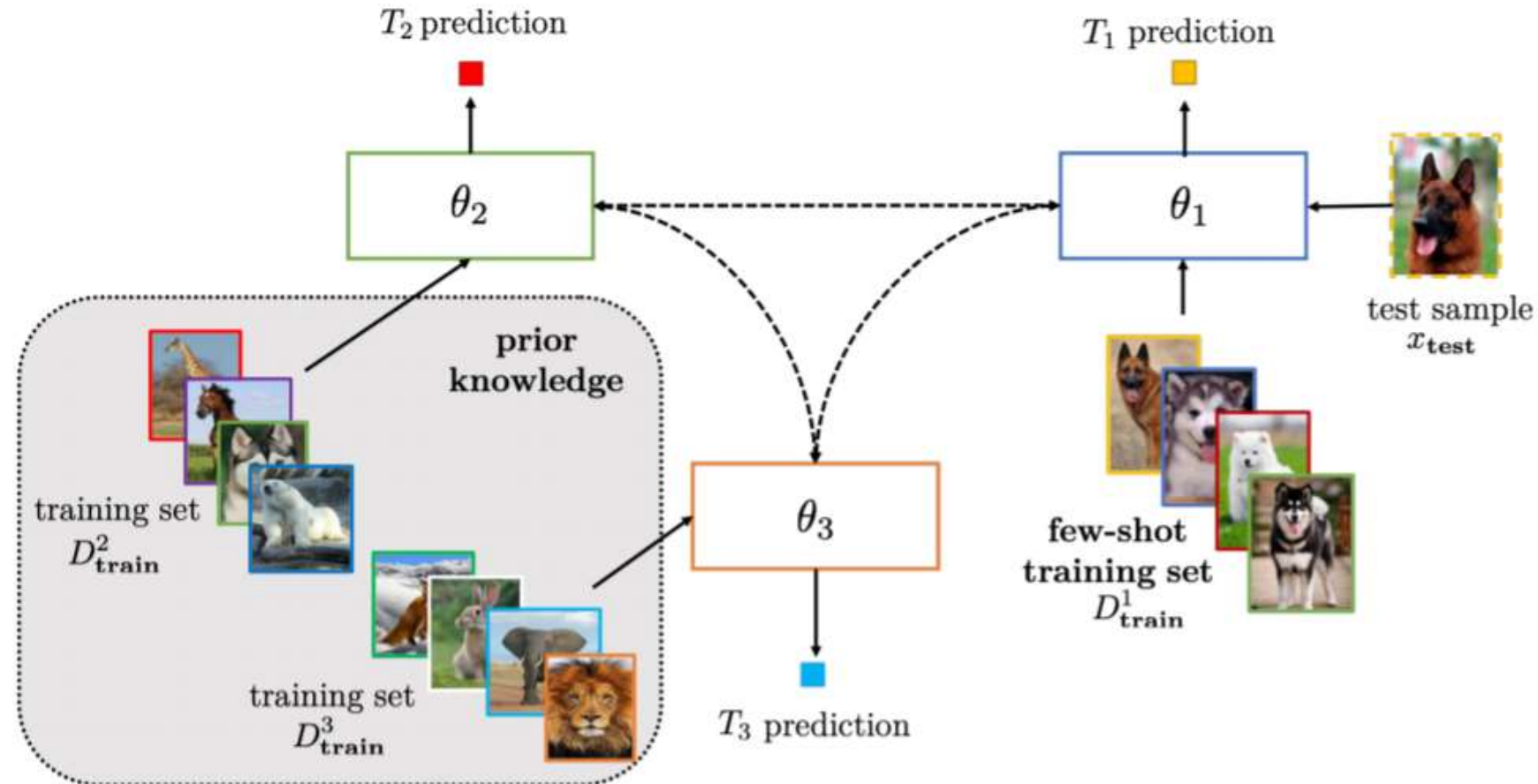
Multitask Learning. Hard parameter sharing

- Имеются общие слои
- Имеются собственные для каждой задачи слои



Multitask Learning. Soft parameter sharing

- Отдельные модели для каждой задачи
- Стремимся к близости весов некоторых слоев



Обучение представлений

Определение

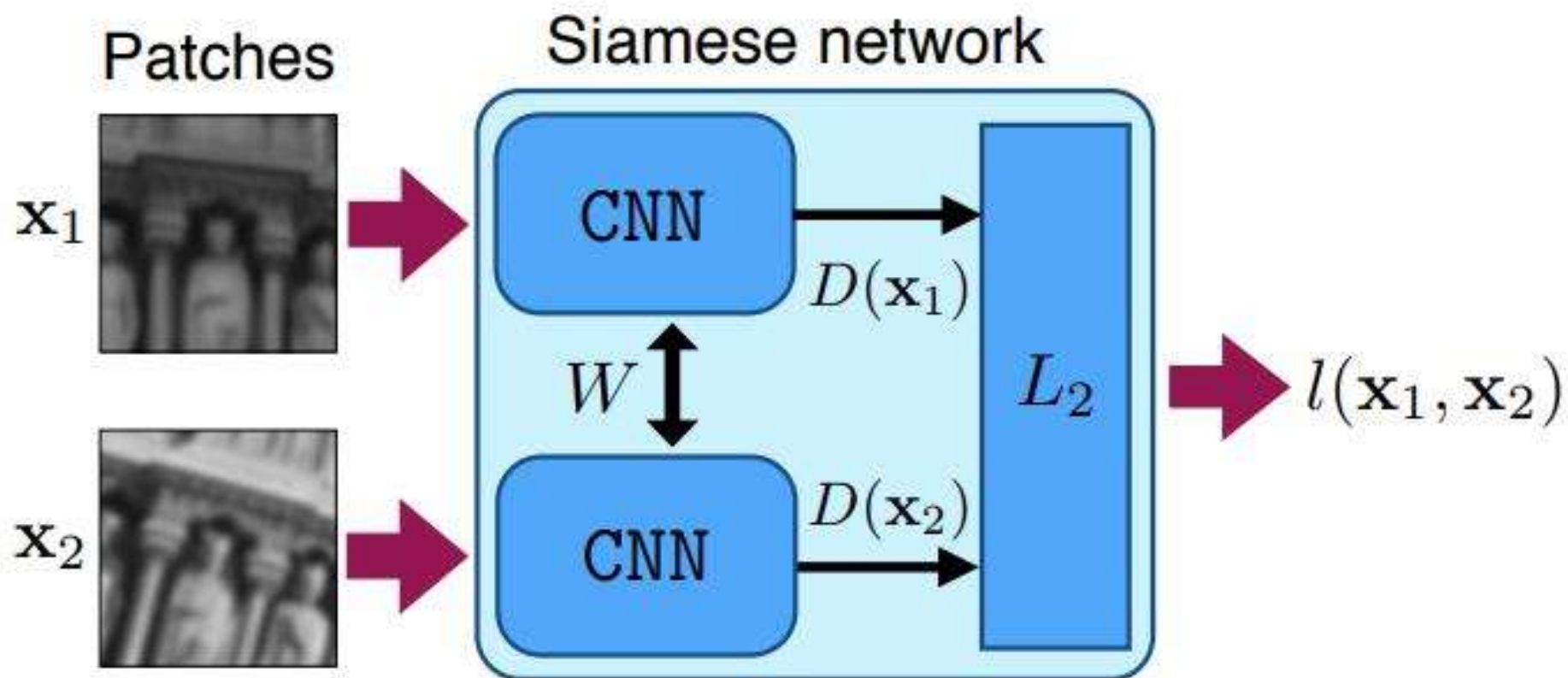
- Метод автоматического получения представлений, необходимых для выявления признаков или классификации данных.

Мотивация

- Для сложных данных таких как, изображение и видео, тяжело алгоритмически выделить важные признаки.

Сиамские сети. Архитектура

- Две CNN
- Одинаковые веса
- Оценивается близость векторов представлений



Сиамские сети. Функция ошибки

- Contrastive loss

$$l(x_i, x_j, z_{ij}) = (1 - z_{ij}) ||h_i - h_j||_2^2 + z_{ij} \max(0, \tau - ||h_i - h_j||_2^2)$$

- Triplet loss

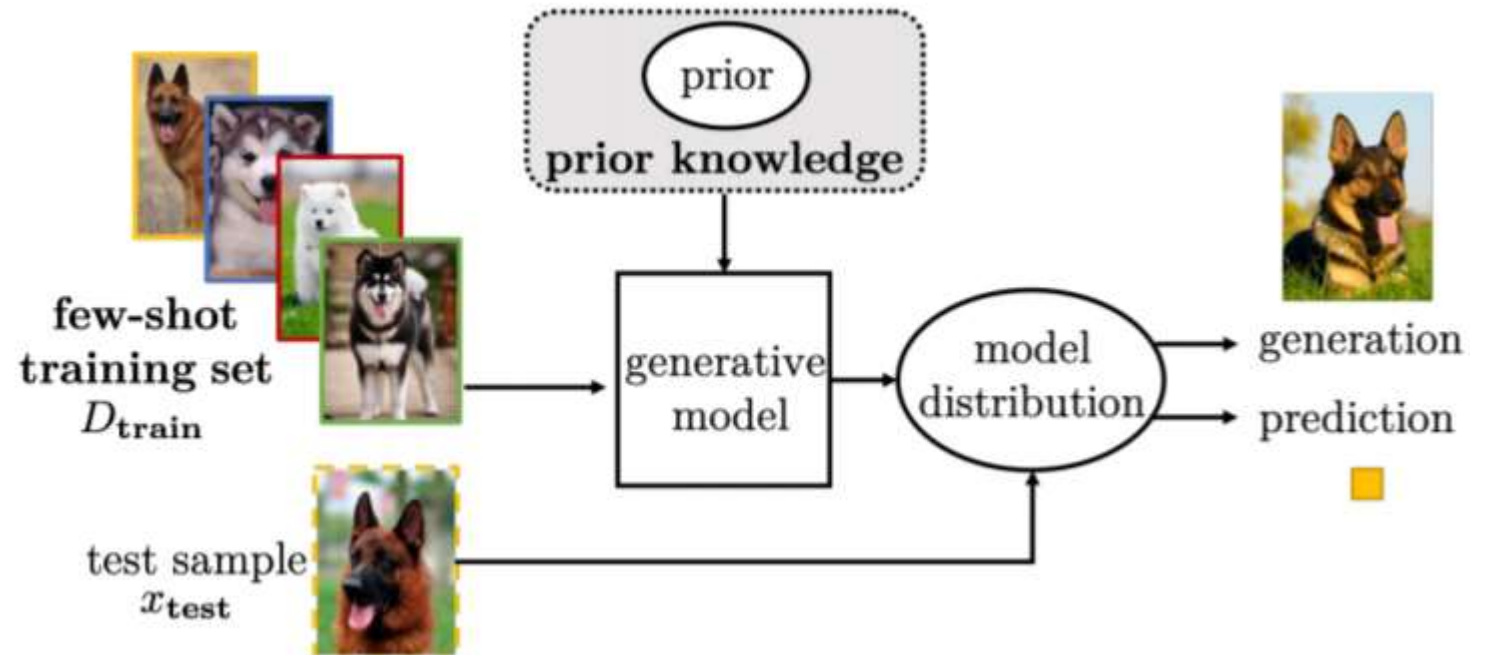
$$l(x_i, x_j, x_k) = \max(0, ||h_i - h_j||_2^2 - ||h_i - h_k||_2^2 + \alpha)$$

Генеративные модели

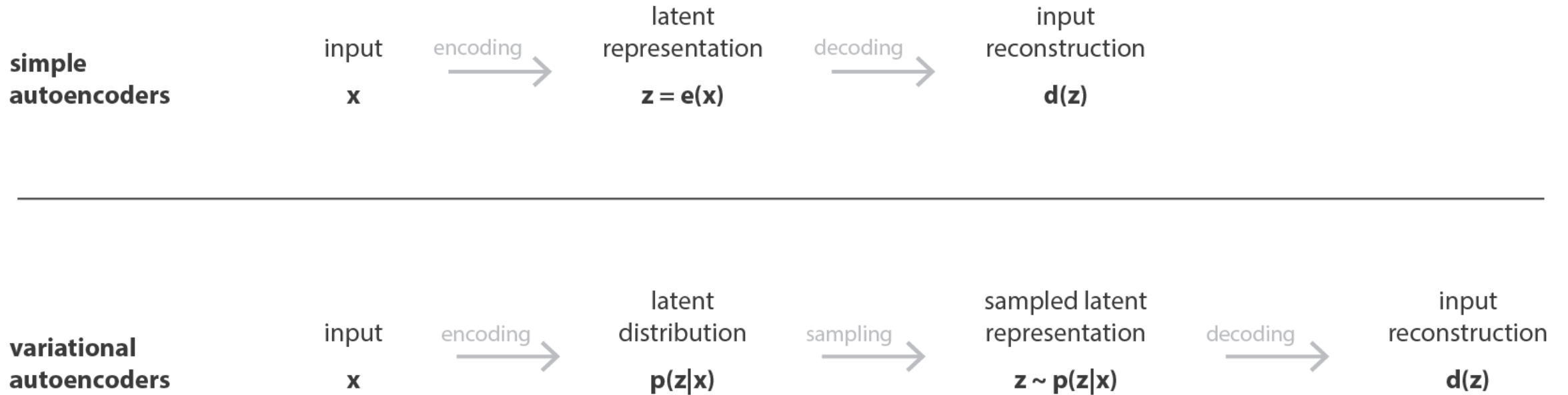
- Используем априорное знание, полученное на похожих данных
- Предсказываем распределение

$$x \sim \int p(x|z; \theta) p(z; \gamma) dz$$

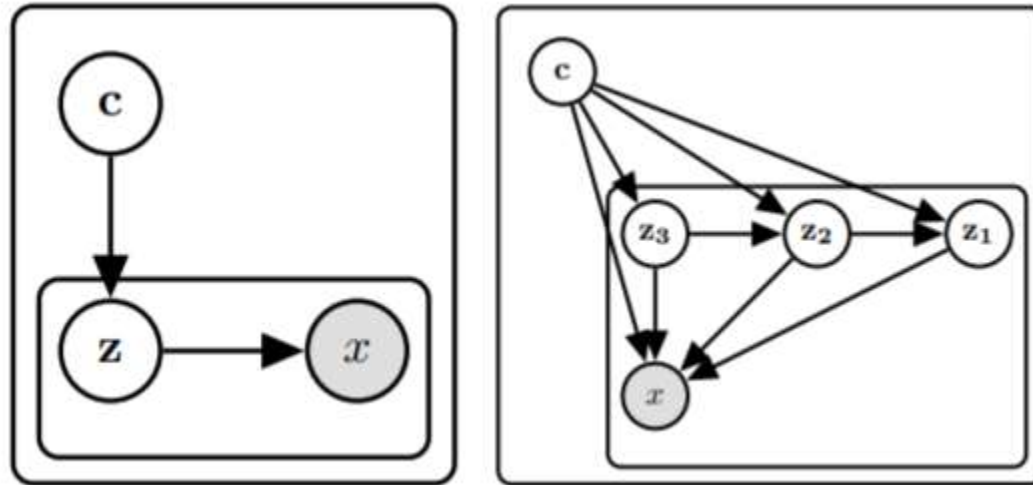
- Априорное знание $p(z; \gamma)$
- Обновляем распределение для наших данных



Variational autoencoder



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

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

Neural Statistician

Плюсы:

- Не требуется много размеченных данных
- Хорошо работает при малом количестве классов
- Высокая обобщающая способность

Минусы:

- Каждый класс рассматривается отдельно
- С увеличением количества классов работает хуже
- Требуется много данных для априорного знания

Test Dataset	Task		Method			
	K Shot	K Way	Siamese	MANN	Matching	Ours
MNIST	1	10	70	-	72	78.6
MNIST	5	10	-	-	-	93.2
OMNIGLOT	1	5	97.3	82.8	98.1	98.1
OMNIGLOT	5	5	98.4	94.9	98.9	99.5
OMNIGLOT	1	20	88.1	-	93.8	93.2
OMNIGLOT	5	20	97.0	-	98.7	98.1

ИСТОЧНИКИ

- <https://arxiv.org/pdf/1904.05046.pdf> FSL
- <https://arxiv.org/ftp/arxiv/papers/1911/1911.07702.pdf>
Сиамские сети
- <https://arxiv.org/ftp/arxiv/papers/1906/1906.09513.pdf>
Сиамские сети
- <https://arxiv.org/pdf/1606.02185v2.pdf> Neural Statistician