

# **Эпистемологическое значение генерализации в машинном обучении**

Титов Сергей

# План доклада

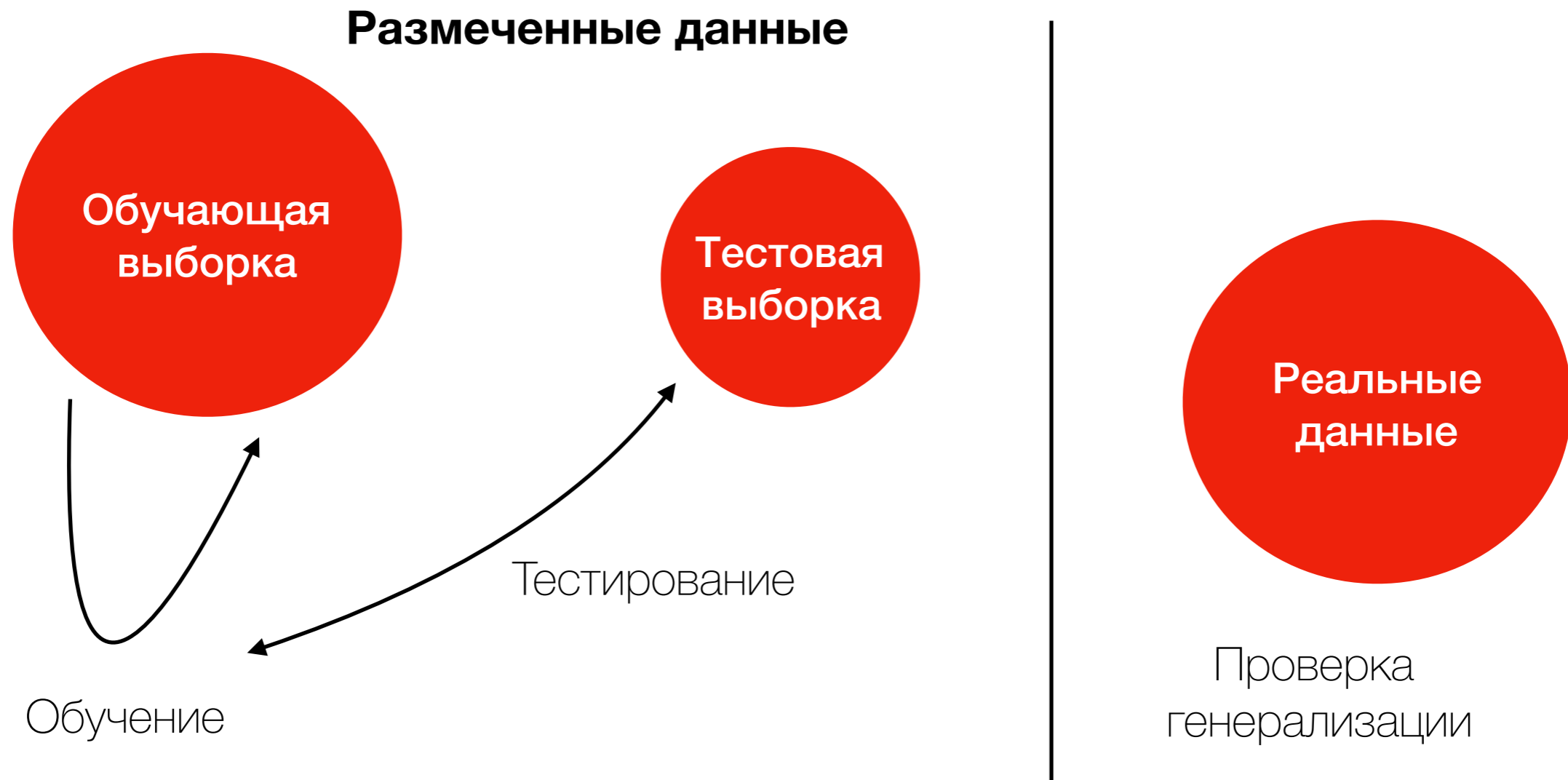
- Основные концепции генерализации в машинном обучении
- Современные исследования генерализации
- Возможные эпистемологические подходы к решению проблем

# Основные концепции машинного обучения

Выделяют несколько основных типов обучения:

1. Обучение с учителем
2. Обучение без учителя
3. Обучение с подкреплением
4. Обучение с частичным привлечением учителя
5. Бесовские методы
6. Методы бустинга

# Основные концепции машинного обучения



# Основные концепции машинного обучения

**Генерализация - способность алгоритма обобщать “выученные” на данных зависимости на новых, ранее им не виденные данные.**

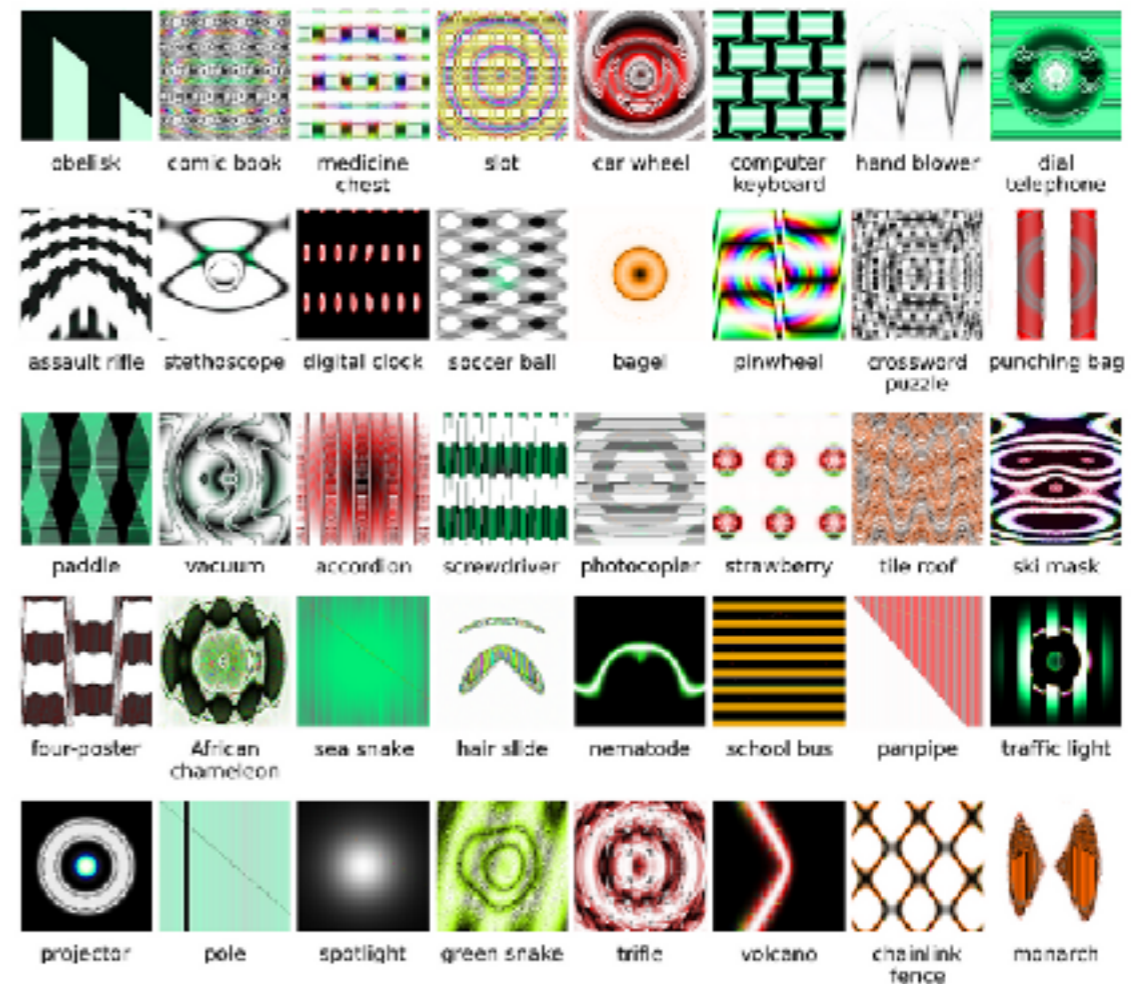
**Обратный эффект от генерализации - переобучение алгоритма. В этом случае он “запоминает” все предоставленные варианты.**

# Проблема

**На данный момент не существует точного объяснения почему генерализация происходит. Более того - при одинаковых настройках алгоритма она может принимать различные значения.**

# Современные исследования

Одно из важных направлений современных исследований в науке о данных - попытки понять и представить как именно происходит генерализация. Одна из самых показательных работы в этой области - Нгуена и коллег, где они продемонстрировали “идеальные” классы для нейронной сети по распознаванию образов



# Современные исследования

В более современных работах основным методом является искажение оригинальных данных. Основная аргументация заключается в том, что:

1. Искажение данных - не меняет проблему обучения (оптимизации)
2. На искаженных данных алгоритм все равно завершает процесс обучения, запоминая все предъявленные примеры
3. Таким образом, остается неизвестным почему, алгоритм обладая возможностью запомнить данные - в некоторых случаях генерализуется



# Современные исследования

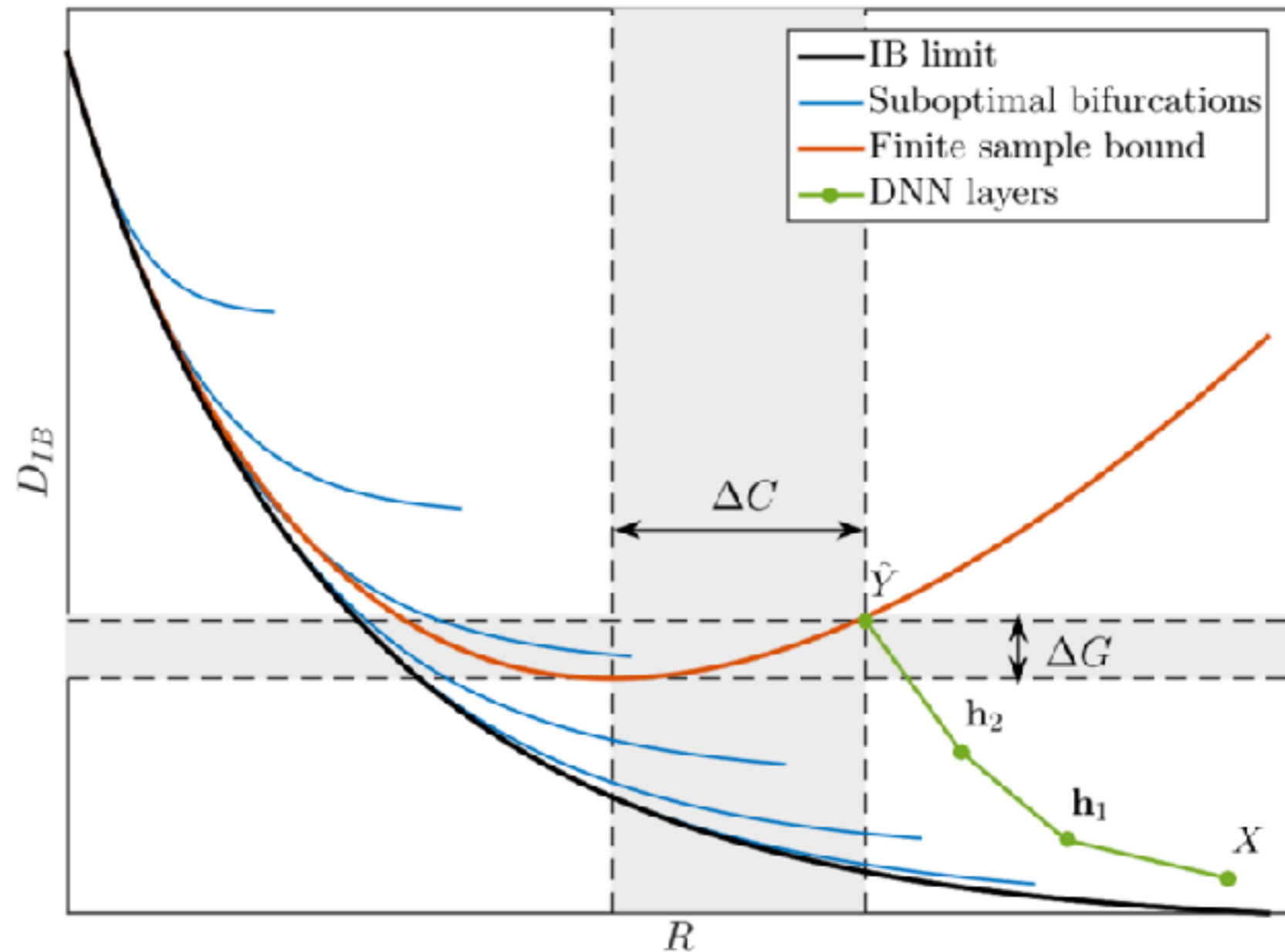
Для того, чтобы пролить свет на процесс генерализации проводится ее изучение с точки зрения теории информации.

Для работы в этом направлении Нафтали Тишби предлагает концепцию “информационного бутылочного горлышка” в рамках которой он предлагает изучать отношения между оптимальной и реальной репрезентации совместной информации о явлении.

# Современные исследования

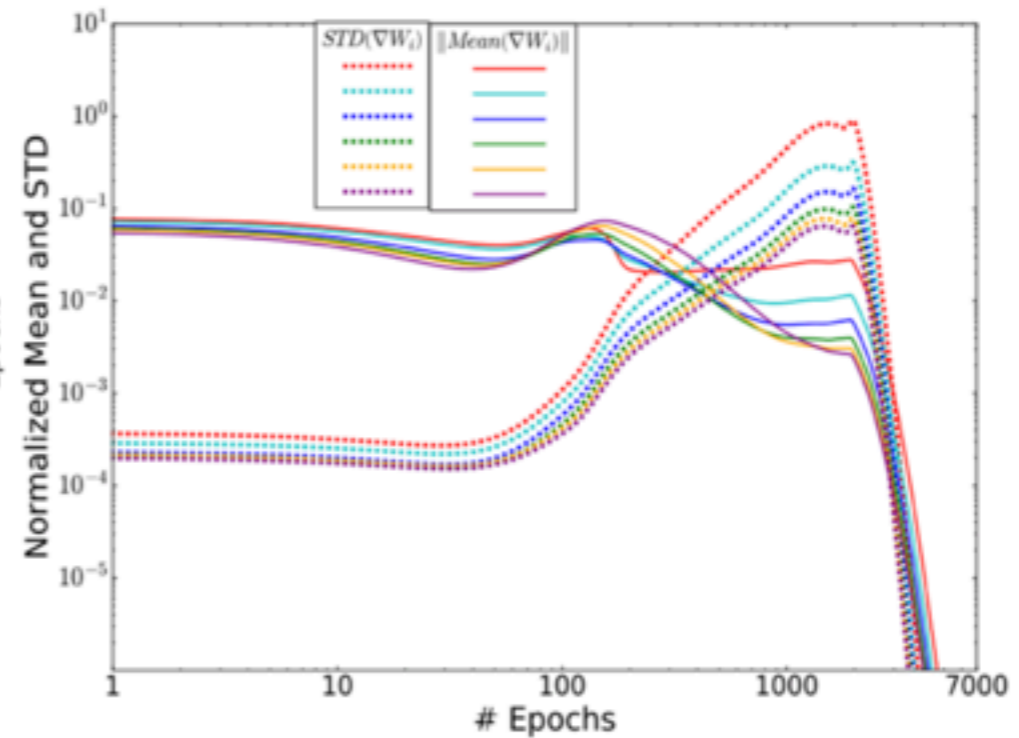
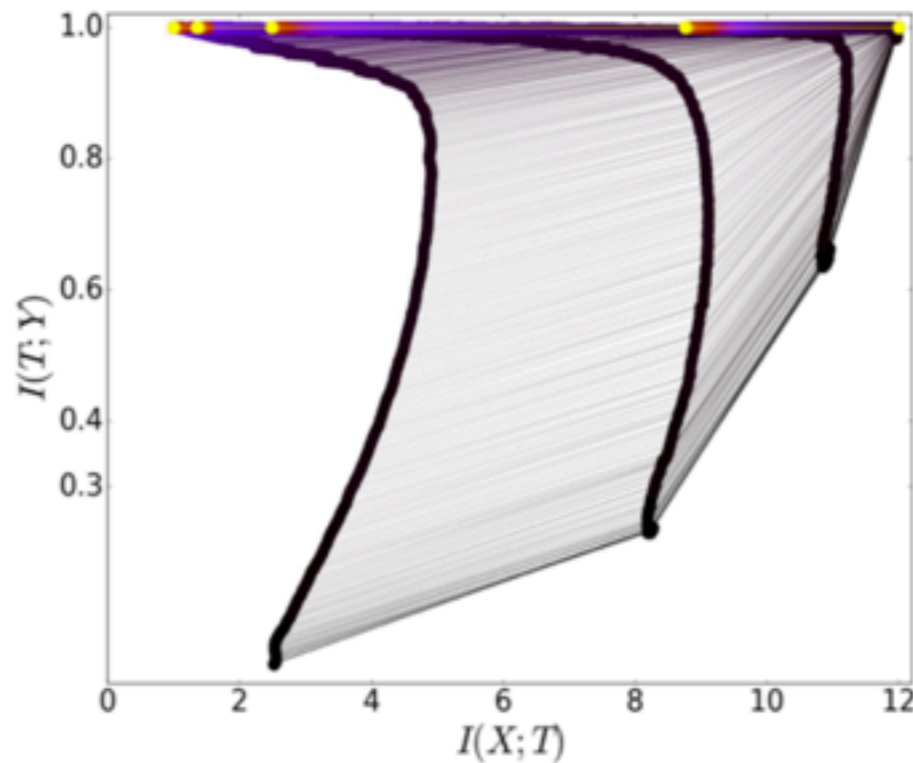
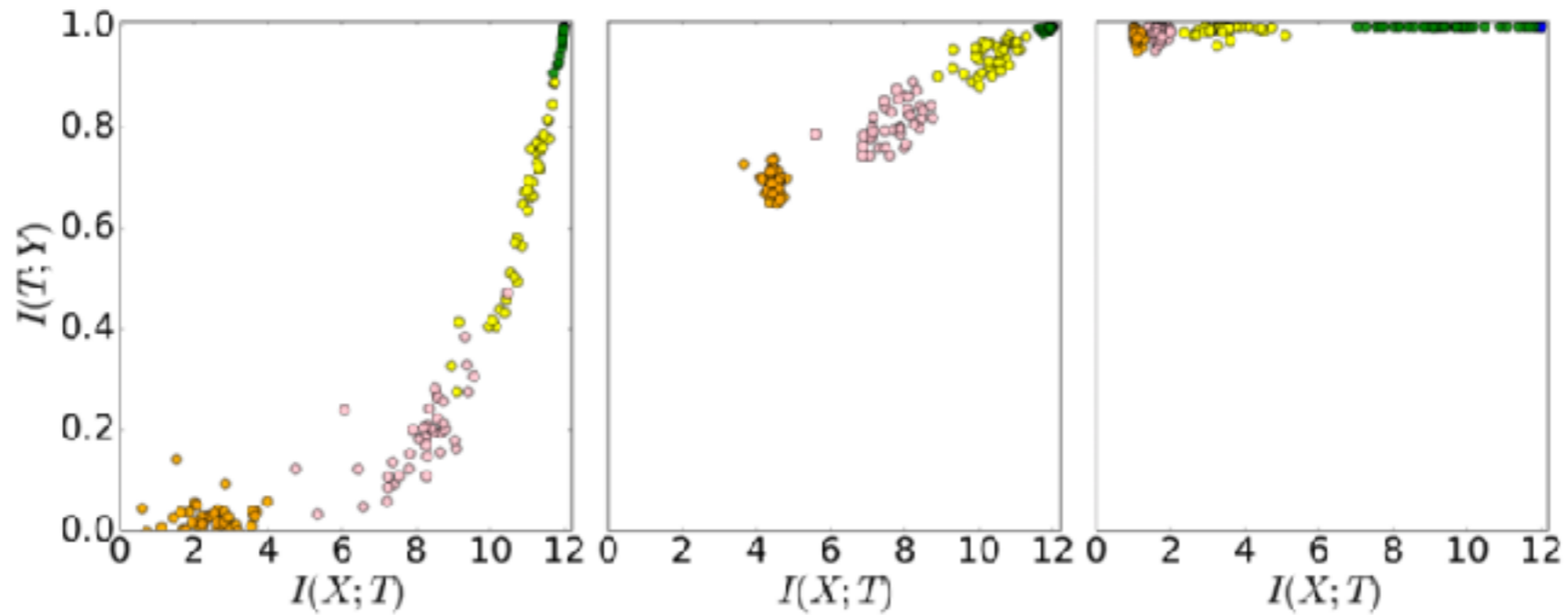
$D$  – информационное расхождение (информация не учитываемая  $X()$ )

$R$  – сложность репрезентации феномена



$$d_{IB}(X, \hat{X}) = D[p(Y|X)] ||| [p(Y|\hat{X})]$$

# Современные исследования



**Спасибо за внимание**

