Исследование способов согласования моделей с помощью снижения размерности пространства

Яушев Фарух Рамильевич

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Консультант Р. Исаченко
Научный руководитель д.ф.-м.н. В. В. Стрижов

Москва
2020 г

Цель работы

Задача

Построить модель прогнозирования целевой переменной в пространстве высокой размерности

Проблема

Исходные данные имеют избыточную размерность. Зависимости в пространствах целевои и независимои переменных приводит к неустоичивости модели.

Метод решения

Построить модель, учитывающую структуру пространств независимой и целевой переменных с согласованием скрытых представлений.

Данные

Дана выборка $(\mathbf{X}, \mathbf{Y})$, где $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]^T \in \mathbb{R}^{n \times m}$ — матрица независимых переменных, $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n]^T \in \mathbb{R}^{n \times k}$ — матрица целевых переменных.

Предполагается, что между $\mathbf{X}$ и целевой $\mathbf{Y}$ существует зависимость

$$\mathbf{Y} = f(\mathbf{X}) + \boldsymbol{\varepsilon},$$

где $f : \mathbb{R}^{n \times m} \to \mathbb{R}^{n \times k}$ — функция регрессионной зависимости, $\boldsymbol{\varepsilon}$ - матрица регрессионных ошибок.

В качестве функции ошибки используется

Необходимо воостановить зависимость f по заданной выборке.

Базовый алгоритм

В качетсве базового алогоритма используется PLS (Projection to Latent Space)

PLS — алгоритм для восстановления связи между двумя наборами данных $\mathbf{X}$ и $\mathbf{Y}$. Алгоритм проецирует $\mathbf{X}$ и $\mathbf{Y}$ на латентное пространство $\mathbb{R}^p$ меньшей размерности. PLS находит матрицы исходных данных $\mathbf{X}$ и $\mathbf{Y}$ в латентном пространстве $\mathbf{T}$ и $\mathbf{U}$ соответственно. Матрица объектов $\mathbf{X}$ и целевая матрица $\mathbf{Y}$ проецируются на латентное пространство следующим образом:

$$\underset{n\times m}{\mathbf{X}} = \underset{n\times p}{\mathbf{T}} \cdot \underset{p\times m}{\mathbf{P}} + \underset{n\times m}{\mathbf{F}}, \tag{1}$$

$$\underset{n\times k}{\mathbf{Y}} = \underset{n\times p}{\mathbf{U}} \cdot \underset{p\times k}{\mathbf{Q}} + \underset{n\times k}{\mathbf{E}}, \tag{2}$$

где $\mathbf{T}$ и $\mathbf{U}$ — матрицы описания объектов и исходов в латентном пространстве; $\mathbf{P}$ и $\mathbf{Q}$ — матрицы перехода из латентного пространства в исходное; $\mathbf{F}$, $\mathbf{E}$ — матрицы остатков.

Определение 1

Параметрическая функция $\varphi_1 : \mathbb{R}^{n \times m} \to \mathbb{R}^{n \times p}$, переводящая исходные данные в латентное пространство, называется **функцией кодирования**.

Определение 2

Параметрическая функция $\varphi_2 : \mathbb{R}^{n \times k} \to \mathbb{R}^{n \times p}$, переводящая данные из латентного пространства в исходное, называется **функцией восстановления**.

Определение 3

Функция $g : \mathbb{R}^{n \times p} \times \mathbb{R}^{n \times p} \to \mathbb{R}$, связывающая два низкоразмерных латентных представления, называется **функцией согласования**.

Постановка задачи

Общая схема

$$
\begin{tikzcd}
\underset{n\times m}{\mathbf{X}} \arrow[rr, "f"] \arrow[d, bend right, "\varphi_1"'] & & \underset{n\times k}{\mathbf{Y}} \arrow[d, bend left, "\psi_1"] \\
\underset{n\times p}{\mathbf{T}} \arrow[u, "\varphi_2"'] \arrow[rr, leftrightarrow, "g"] & & \underset{n\times p}{\mathbf{U}} \arrow[u, "\psi_2"]
\end{tikzcd}
$$

Оптимальные параметры $\theta^*_{\varphi_1}, \theta^*_{\psi_1}$ для функций кодирования φ_1 и ψ_1 находятся из следующей задачи параметрической оптимизации:

$$(\theta^*_{\varphi_1}, \theta^*_{\psi_1}) = \underset{(\theta_{\varphi_1}, \theta_{\psi_1})}{\arg\max}\left[g\left(\varphi_1(\mathbf{X}; \theta_{\varphi_1}), \psi_1(\mathbf{Y}; \theta_{\psi_1})\right)\right]. \tag{3}$$

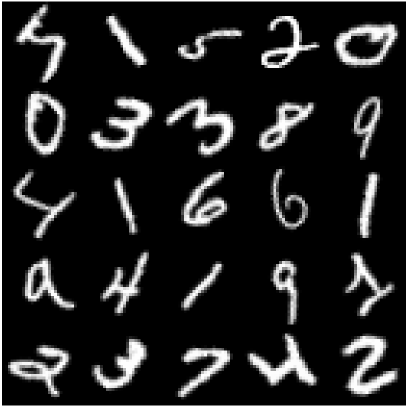
Постановка задачи

После перехода в латентное пространство между $\mathbf{T}$ и $\mathbf{U}$ существует зависимость
$$\mathbf{U} = h(\mathbf{T}) + \boldsymbol{\eta},$$
где $h : \mathbb{R}^{n \times p} \to \mathbb{R}^{n \times p}$ — функция регрессионной зависимости, $\boldsymbol{\eta}$ — матрица регрессионнх ошибок.

Оптимальные параметры для h выбираются минимизацией функции ошибки. Используем квадратичную функцию ошибки потерь $\mathcal{L}$ на $\mathbf{T}$ и $\mathbf{U}$:
$$\mathcal{L}(h|\mathbf{T}, \mathbf{U}) = \Bigg\| \underset{n \times p}{\mathbf{U}} - h(\underset{m \times p}{\mathbf{T}}) \Bigg\|_2^2 \to \min_h .$$

Финальная прогностическая модель имеет вид:
$$f = \psi_2 \circ h \circ \varphi_1.$$

Для того, чтобы показать, что игнорирование нелинейных зависимостей может привести к неудовлетворительным результатам проведем сравнение DeepCCA и CCA на задаче классификации зашумленных цифровых изображений.



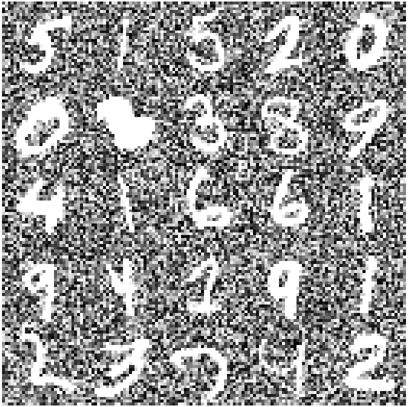


Рис.: Зашумленные изображений из набора данных MNIST

Канонический анализ корреляций (CCA)

Функция согласования:

$$g(\mathbf{X}\mathbf{w}_\mathbf{x}, \mathbf{Y}\mathbf{w}_\mathbf{y}) = \mathbf{corr}(\mathbf{X}\mathbf{w}_\mathbf{x}, \mathbf{Y}\mathbf{w}_\mathbf{y}),$$

Функции кодирования:

$$\varphi_1(\mathbf{X}) = \mathbf{W}_\mathbf{x}^\mathbf{T}\mathbf{X}, \quad \psi_1(\mathbf{Y}) = \mathbf{W}_\mathbf{y}^\mathbf{T}\mathbf{Y},$$

Таблица: Получение нового признакового пространство размерности 15 с использованием DeepCCA и CCA. Показателем эффективности будет точность классификации линейного SVM (ACC).

	DeepCCA(L=3)	CCA
Validation data	92.74%	76.21%
Test data	92.14%	76.07%

Решается задача восстановления правой части изображения по левой.

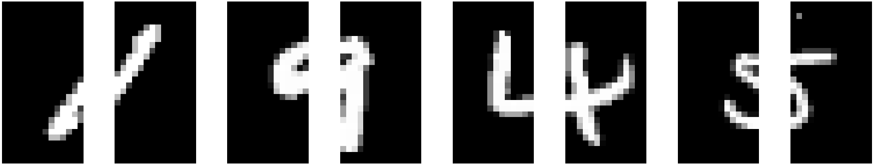


Рис.: Набор данных MNIST, каждое изображение в котором разделили пополам.

Таблица: Восстановление правой части изображения по левой с использованием различных моделей. Для измерения качества моделей считается среднеквадратическое отклонения от оригинального изображения.

	EncNet1	LinNet1	EncNet2	LinNet2	DumbNet	PLS
Кол-во весов	283k	239k	283k	239k	283k	-
MSE loss	0.147	0.235	0.149	0.236	0.128	0.188

Эксперимент