

Практикум по основам машинного обучения, 2016

Вахитов А.Т.

February 19, 2016

1 Зависимость результата обучения от размера выборки и сложности множества гипотез

Берем NIST Handwritten digits dataset.

Составляем программу, осуществляющую классификацию цифр 1 и 5 по вектору, составленному из значений пикселей изображения заданного размера, с помощью линейной регрессии с константным сдвигом.

	Целевое разрешение	N=200	N = 2000	N = 20000
В итоге, заполняем таблицу:	4×4	E_{in}/E_{out}
	8×8
	16×16

Строим графики E_{in} , E_{out} для каждого разрешения (всего получится 6 графиков) в осях E , N , где E - соответствующая вероятность ошибки. Считаем E_{out} по всему тестовому набору, обучение производим по N случайно выбранным примерам тренировочного набора.