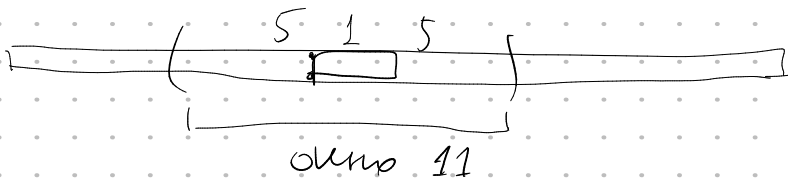


клетчатый шаг по контексту



в этой задаче контекст - граница вектор
 → гистерезис фгиперность

более простое метрику преобразования слов в вектора. Word2Vec

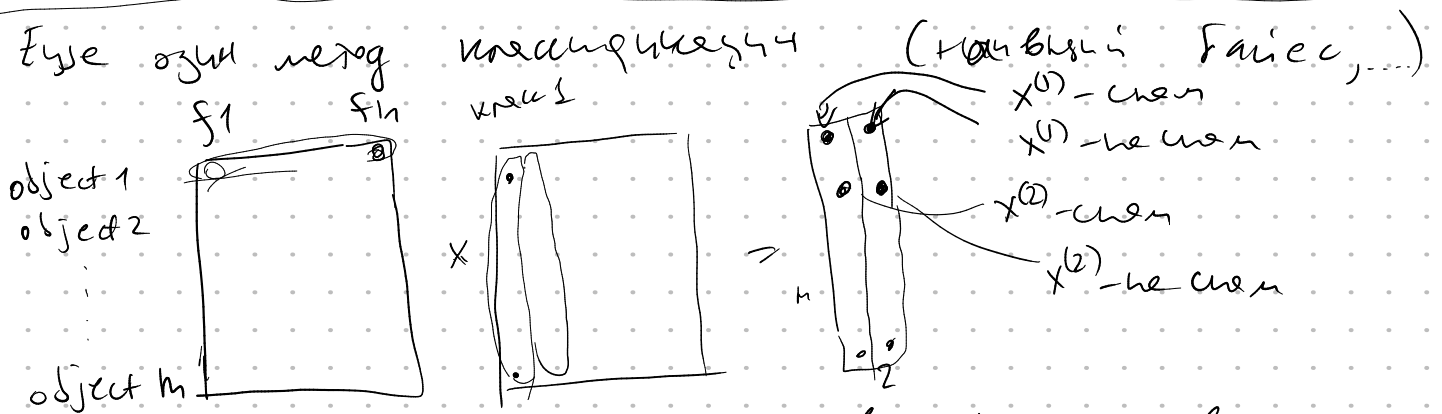
слово → 50-300 значений

$$wzc(king) - wzc(man) \approx wzc(queen) - wzc(woman)$$

$$wzc(aug) - wzc(men) + wzc(woman) - wzc(queen)$$

сенси фгиперность

$$wzc(Russia) - wzc(moscow) \approx wzc(France) - wzc(Paris)$$



признаки $m \times n$

вектор x

вектор признаков $x^{(1)}$

$$x_1^{(1)} \cdot w_1^{spam} + x_2^{(1)} \cdot w_2^{spam} + \dots + x_n^{(1)} \cdot w_n^{spam}$$

вектор признаков $x^{(1)}$

вектор признаков $x^{(2)}$

вектор признаков $x^{(1)}$

вектор признаков $x^{(2)}$

вектор признаков $x^{(1)}$

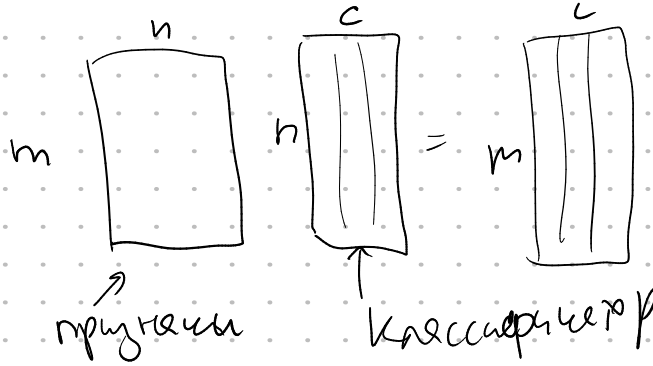
вектор признаков $x^{(2)}$

вектор признаков $x^{(1)}$

вектор признаков $x^{(2)}$

вектор признаков не spam: $x_1^{(1)} \cdot w_1^{he} + x_2^{(1)} \cdot w_2^{he} + \dots + x_n^{(1)} \cdot w_n^{he}$

что такое

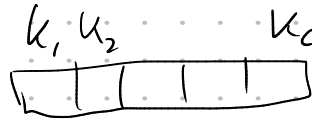


c-классов

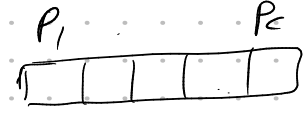
базис для

столбцов

объект - базис



sm



хотите иметь не базис, а вероятности.

$$p_i \geq 0$$

$$\sum p_i = 1$$

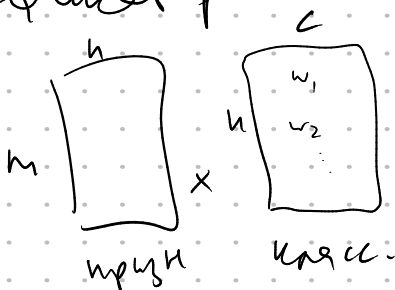
soft-max =

$$p_i = \frac{e^{k_i}}{e^{k_1} + \dots + e^{k_c}}$$

$$\text{soft-max}([1, -1, 2]) =$$

$$\left[\frac{e^1}{e^1 + e^{-1} + e^2}, \frac{e^{-1}}{e^1 + e^{-1} + e^2}, \frac{e^2}{e^1 + e^{-1} + e^2} \right]$$

Кросс-энтропия

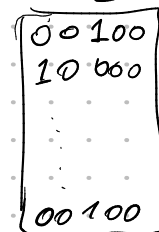


softmax
но
спрощен



\bar{Y}
результат

c-классов



Y
прав.

прав. ответы

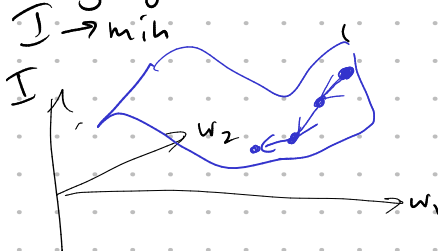
лучше, чем в naive bayes.

Алгоритм минимизации: $I = \sum_{i=1}^m (\bar{Y}_i - Y_i)^2 \rightarrow \min$.

Задача

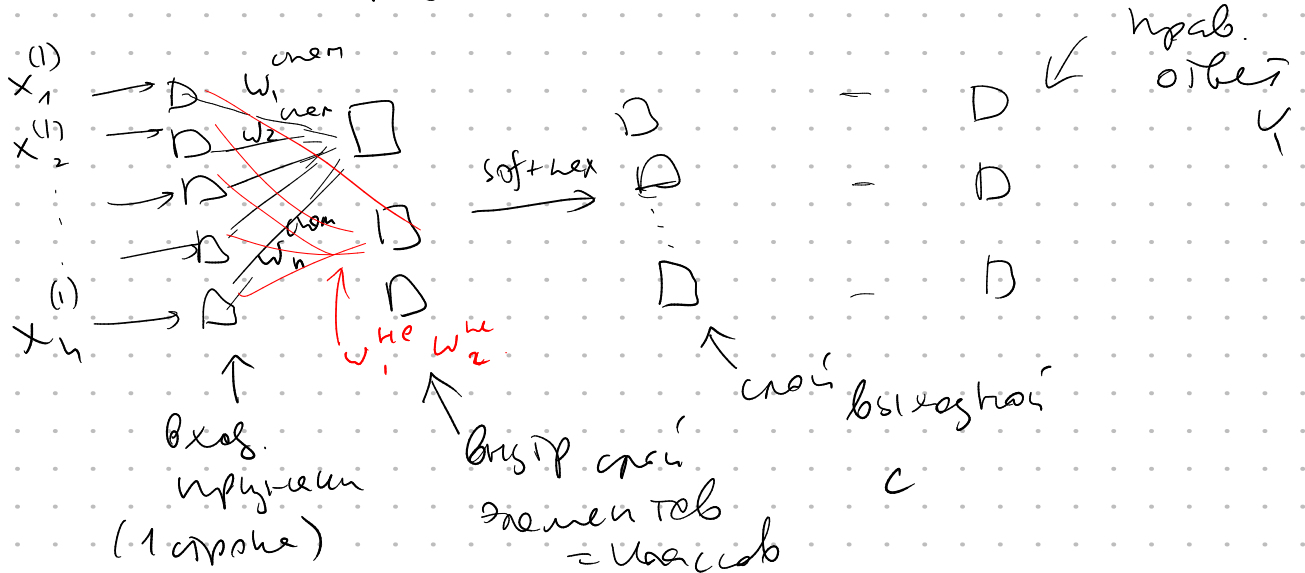
оптимизации.

В нейронной сети минимизация \rightarrow градиентный спуск



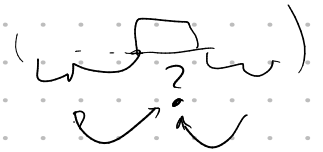
$$\nabla I = \left[\frac{\partial I}{\partial w_1}, \frac{\partial I}{\partial w_2} \right]$$

на-прямую изобразим схему классиф.

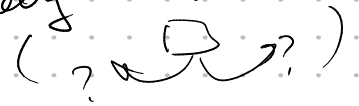


это классификация с помощью нейр. сети.

Угест. предполагает 1) слово по контексту cbow



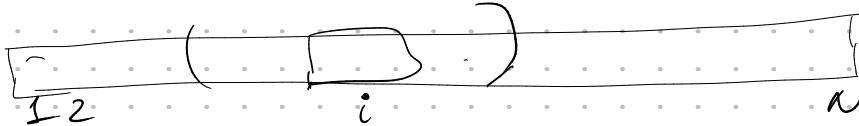
2) контекст по слову skip-gram



слова слева \equiv контекст слова

нахождение контекста \Leftrightarrow нахождение слова
(дистрибутивная семантика)

cbow = continuous bag of words / skip-gram



корпус

примеры для обучения N итер. или словосочетания

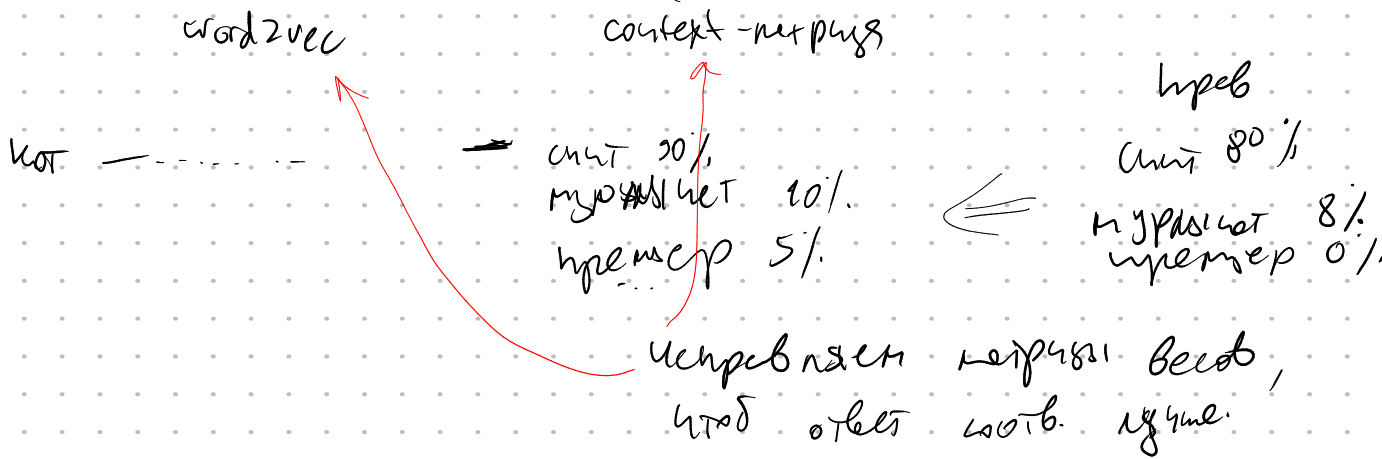
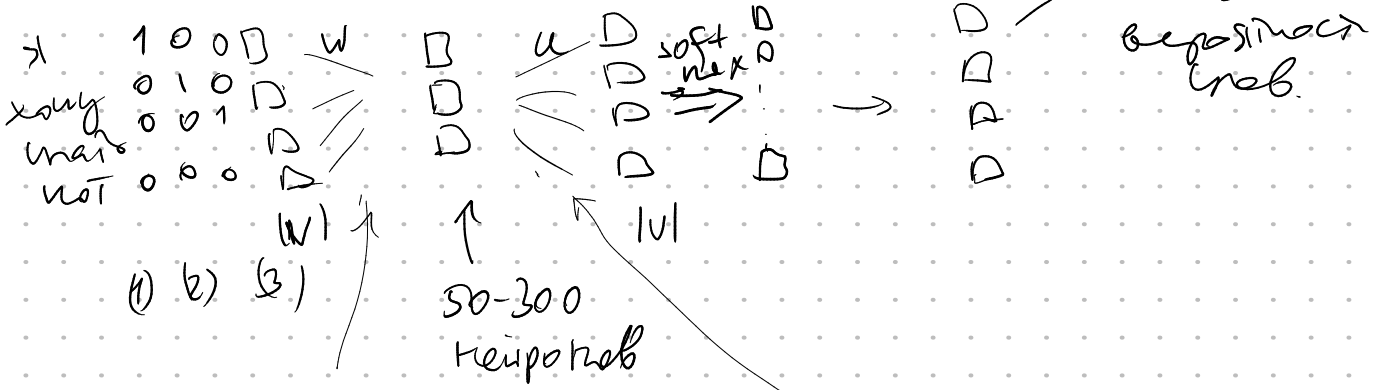
в корпусе

w_i

0	▷
0	▷
0	▷
1	▷
0	▷
0	▷
0	▷
0	▷

признаки слова: (для слова, 0 для ост.)

9 xony che. 76



У 50-300 нейронів можна розв'язати задачу. Оскільки контекст
 Все нахороше виводить - то є його слово word2vec
 потрібне

