Жеглов Дмитрий Андреевич, 4 домашняя работа

Задание 1.

$\boxed{3.}$ $\prod_{i=1}^{\ell} \left(\frac{1}{\exp(-\langle w, x_i\rangle y_i)+1}\right) \to \max$ Функция правдоподобия

возьмем ln т.к. монотонная функция

$$\Rightarrow \sum_{i=1}^{\ell} \cancelto{0}{\ln 1} - \sum_{i=1}^{\ell} \ln\left(1+\exp(-\langle w, x_i\rangle y_i)\right) \to \max_w$$

$$\Rightarrow \sum_{i=1}^{\ell} \ln\left(1+\exp(-\langle w, x_i\rangle y_i)\right) \to \min_w$$

получили ~~логистическую~~ логарифмическую функцию потерь

Задание 2.

3 2 $w_{k+1} = w_k - \eta \cdot \mathrm{grad}\, F$, $\eta > 0$

$$\mathrm{grad}\, F_j = \left(\sum_{i=1}^{\ell} \frac{-\exp(-\langle w, x_i\rangle y_i)\, x_i^j y_i}{1+\exp(-\langle w, x_i\rangle y_i)}\right) =$$

$$= \left(-\sum_{i=1}^{\ell} \frac{x_i^j \cdot y_i}{1+\exp(\langle w, x_i\rangle y_i)}\right)$$ при градиентном спуске

Т.к. выборка линейно разделима

$\Rightarrow \exists w : \langle w, x_i\rangle \cdot y_i > 0 \ \forall i \Rightarrow$

$\Rightarrow \forall c > 1, \forall i,\ c\langle w, x_i\rangle y_i > 0 \Rightarrow \langle c\cdot w, x_i\rangle y_i > 0 \Rightarrow$

$c\cdot w$ тоже разделяющий вектор весов

Ф-ция правдоподобия имеет вид из Задачи 1

$$-Q = -\sum_{i=1}^{\ell} \ln\left(1+\exp(-\langle w, x_i\rangle y_i)\right) \to \max_w$$

Устремим $c \to \infty$, получим

$$-Q = -\sum_{i=1}^{\ell} \ln(1+0) \to 0,$$ т.к. $\exp(-\langle c\cdot w, x_i\rangle y_i) \to 0$

т.к. $c\cdot\langle w, x\rangle y_i > 0$ при $c \to \infty$ стремится к бесконечности при любом i

$\boxed{\text{Но } -Q < 0}$ на этих векторах

$\Rightarrow$ вектора весов для максимизации ф-ции правдоподобия не существует, т.к. максимум не достигается

Проблема заключается в неограниченности вектора весов w, попробуем улучшить модель добавив в нее регуляризацию, чтобы штрафовать большие веса

] w имеет распределение, тогда функция правдоподобия имеет вид

$$L = \prod_{i=1}^{\ell} p(x_i, y_i \mid w) \cdot p(w) = \prod_{i=1}^{\ell} p(y_i \mid w, x_i) \cdot p(x_i) \cdot p(w)$$

Т.к. множество классов Y дискретно

Будем максимизировать $L \to \max\limits_{w}$, то же самое, что

$$\ln\left[\prod_{i=1}^{\ell} p(y_i \mid w, x_i)\, p(x_i)\, p(w)\right] \to \max_{w}$$

$$\sum_{i=1}^{\ell} \ln p(y_i \mid w, x_i) + \underbrace{\cancel{\sum_{i=1}^{\ell} \ln p(x_i)}}_{\text{не зависит от } w} + \ln p(w) \to \max_{w}$$

$$p(y_i \mid w, x_i) = \frac{1}{1 + \exp(-\langle w, x_i \rangle y_i)}$$

$$] \; p(w) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\|w\|^2\right) \; [\text{Гауссово}]$$

т.к. не зависит от w

$$\Rightarrow \cancel{\sum_{i=1}^{\ell} \ln 1} - \sum_{i=1}^{\ell} \ln(1 + \exp(-\langle w, x_i \rangle y_i)) + \cancel{\ln\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}} - \frac{1}{2\sigma^2}\|w\|^2 \to \max_{w}$$

$$\Rightarrow \boxed{\sum_{i=1}^{\ell} \ln(1 + \exp(-\langle w, x_i \rangle y_i)) + \frac{1}{2\sigma^2}\|w\|^2 \to \min_{w}}$$

Теперь проблема решения

$$w_{k+1} = w_k + \eta\left[\sum_{i=1}^{\ell} \frac{x_i y_i}{(1 + \exp(\langle w, x_i \rangle y_i)} - \frac{w_k}{\sigma^2}\right]$$

как и в первом случае только с производной регуляризационной составляющей

Задание 3

3.3 При построении ROC кривой все объекты сортируются по оценке классификатора.

Начинаем из начала координат и идем от большей оценки к меньшей. Если текущий объект класса „+1" $\Rightarrow$ увеличиваем TPR т.е. кривая сдвигается вверх на $\frac{1}{\ell_+}$, если у текущего объекта класс „0", то алгоритм допускает на 1 ошибку больше, чем предыдущий $\Rightarrow$ кривая сдвигается вправо на $\frac{1}{\ell_-}$, а к AUC надо прибавить

$$\frac{1}{\ell_+\ell_-}\sum_{j=i+1}^{\ell}[y_{(j)}=1]$$

$$\Rightarrow AUC = \frac{1}{\ell_+\ell_-}\sum_{i=1}^{\ell}[y_{(i)}=0]\sum_{j=i+1}^{\ell}[y_{(j)}=1] =$$

$$= \frac{1}{\ell_+\ell_-}\sum_{i=1}^{\ell}\sum_{j=i+1}^{\ell}[y_{(i)}<y_{(j)}] = \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}(1-[y_{(i)}>y_{(j)}]-[y_{(i)}=y_{(j)}]) =$$

$$= \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}(1-[y_{(i)}=y_{(j)}]) - \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}[y_{(i)}>y_{(j)}] =$$

$$= \frac{\ell(\ell-1)}{2\ell_+\ell_-} - \frac{\ell_+(\ell_+-1)}{2\ell_+\ell_-} - \frac{\ell_-(\ell_--1)}{2\ell_+\ell_-} - \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}[y_{(i)}>y_{(j)}] =$$

$$= 1 - \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}[y_{(i)}>y_{(j)}]$$

$$\Rightarrow DP = \frac{2\ell_-\ell_+}{\ell(\ell-1)}(1-AUC)$$

Задание 4

В4 Нам дан наивный байесовский классификатор

У нас 2 класса 1 и −1, а значит необходимо выполнение равенства:

$$\lambda_{-1} P(y=-1)\, p(x \mid y=-1) = \lambda_1 P(y=1)\, p(x \mid y=1)$$

Т.к. $x = (x_1, x_2)$ вектор признаков и признаки независимы

$$\Rightarrow \lambda_{-1} P(y=-1)\, p(x_1 \mid y=-1)\, p(x_2 \mid y=-1) =$$

$$= \lambda_1 P(y=1)\, p(x_1 \mid y=1)\, p(x_2 \mid y=1)$$

Т.к. признаки Гауссовы, т.е. $p(x_i \mid y=s) = N(a_{is}, \sigma_{is}^2)$

$s = -1, 1$

тогда по выборке оценим параметры

1) Первая выборка, по картинке видно, что

$a_{1-1} = \frac{1}{2}$

$a_{11} = \frac{3}{2}$

$a_{2-1} = \frac{3}{2}$

$a_{21} = \frac{1}{2}$

$\sigma_{is}^2 = \sigma^2$ одинаковы для всех. Тогда получим

$$\lambda_{-1} P(y=-1) \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{1}{2\sigma^2}(x_1-\frac{1}{2})^2\right)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(x_2-\frac{3}{2}\right)^2\right) =$$

$$= \lambda_1 P(y=1) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(x_1-\frac{3}{2}\right)^2\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(x_2-\frac{1}{2}\right)^2\right)$$

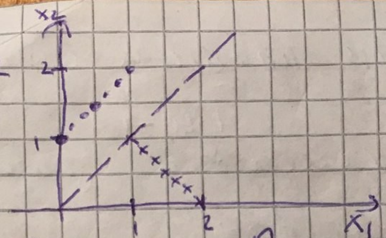
Количество элементов каждого класса одинаковое

$\Rightarrow$ скорее всего $P(y=-1) = P(y=1)$, а так же положим

одинаковый штраф $\lambda_{-1} = \lambda_1$, сократим и применим ln

$$\Rightarrow -\frac{1}{2\sigma^2}\left(x_1-\frac{1}{2}\right)^2 - \frac{1}{2\sigma^2}\left(x_2-\frac{3}{2}\right)^2 = -\frac{1}{2\sigma^2}\left(x_1-\frac{3}{2}\right)^2 - \frac{1}{2\sigma^2}\left(x_2-\frac{1}{2}\right)^2$$

Получим $x_1 = x_2$



2) Рассмотрим вторую выборку. По рисунку оценим парам.

$a_{1,-1} = -1$

$a_{2,-1} = 1$

$a_{11} = 1$

$a_{21} = 2$

$\sigma_{is}^2 = \sigma^2$ для всех

Аналогично первому случаю получим

$$-\frac{1}{2\sigma^2}(x_1+1)^2 - \frac{1}{2\sigma^2}(x_2-1)^2 = -\frac{1}{2\sigma^2}(x_1-1)^2 - \frac{1}{2\sigma^2}(x_2-2)^2$$

Получим $\cancel{4}x_2 = \frac{3}{2} - 2x_1$

Задание 5

$\boxed{35}$ Наивный байесовский классификатор в случае двух классов 0 и 1 а так же бинарных признаков $\xi_j \in \{0,1\}$ имеет вид

$$a(x) = \Big[\ln \lambda_1 \underbrace{P(y=1)}_{=p_1} - \ln \lambda_0 \underbrace{P(y=0)}_{=p_0} + \sum_{j=1}^{n} \big[\ln p_{1j}(\xi_j) - \ln p_{0j}(\xi_j)\big] \Big] \Rightarrow$$

$$\Rightarrow \ln \lambda_1 p_1 - \ln \lambda_0 p_0 + \sum_{j=1}^{n} \Big[(\ln p_{1j}(0) - \ln p_{0j}(0))\, I(\xi_j = 0) + (\ln p_{1j}(1) - \ln p_{0j}(1))\, I(\xi_j = 1) \Big] =$$

$$= \ln\left(\frac{\lambda_1 p_1}{\lambda_0 p_0}\right) + \sum_{j=1}^{n} \Big[\ln p_{1j}(0) - \ln p_{0j}(0) + (\ln p_{1j}(1) - \ln p_{0j}(1) + \ln p_{0j}(0) - \ln p_{1j}(0))\, \xi_j \Big]$$

Положим $\alpha_0 = \ln\left(\frac{\lambda_1 p_1}{\lambda_0 p_0}\right) + \sum_{j=1}^{n} \big[\ln p_{1j}(0) - \ln p_{0j}(0)\big]$

$\alpha_j = \ln p_{1j}(1) - \ln p_{0j}(1) + \ln p_{0j}(0) - \ln p_{1j}(0)$

$j = \overline{1,n}$

$$\Rightarrow a(x) = \Big[\alpha_0 + \sum_{j=1}^{n} \alpha_j \xi_j > 0 \Big] \quad \blacksquare$$

Задание 6

$\boxed{36}$ $X_l = \{\vec{x}_i, y_i\}_{i=1}^{l}$, $\vec{x}_i \in \mathbb{R}^2$, $y_i \in Y = \{0,1\}$

λ_y штраф за ошибку в классе y.

$\ln \lambda_y P(y) = C_y$, $y \in Y = \{0,1\}$

Функции правдоподобия Гауссовы, $\vec{\mu}_0 = \begin{pmatrix} a \\ b \end{pmatrix}$, $\vec{\mu}_1 = \begin{pmatrix} -a \\ -b \end{pmatrix}$

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & S \end{pmatrix}$$

Байесовский классификатор для двух классов

$$a(x) = \left[\ln \frac{\lambda_1 P(y=1)}{\lambda_0 P(y=0)} + \ln p_1(x) - \ln p_0(x) > 0\right] =$$

$$= \left[C_1 - C_0 - \frac{1}{2}(x-\vec{\mu}_1)^T \Sigma^{-1} (x-\vec{\mu}_1) + \ln\frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} + \frac{1}{2}(x-\vec{\mu}_0)^T \Sigma^{-1} (x-\vec{\mu}_0) - \right.$$

$$\left. - \ln\frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} > 0\right] =$$

$$= \left[C_1 - C_0 + \frac{1}{2}(x_1-a, x_2-b)\begin{pmatrix} 1 & 0 \\ 0 & S^{-1} \end{pmatrix}\begin{pmatrix} x_1-a \\ x_2-b \end{pmatrix} - \frac{1}{2}(x_1+a, x_2+b)\begin{pmatrix} 1 & 0 \\ 0 & S^{-1} \end{pmatrix}\begin{pmatrix} x_1+a \\ x_2+b \end{pmatrix} > 0\right]$$

$$= \left[C_1 - C_0 + \frac{1}{2}(x_1-a)^2 + \frac{1}{2}S^{-1}(x_2-b)^2 - \frac{1}{2}(x_1+a)^2 - \frac{1}{2}S^{-1}(x_2+b)^2 > 0\right] =$$

$$= \left[C_1 - C_0 - x_1 \cdot a - S^{-1}x_2 \cdot b - x_1 \cdot a - S^{-1}x_2 \cdot b > 0\right] =$$

$$= \left[C_1 - C_0 - 2ax_1 - 2bS^{-1}x_2 > 0\right]$$

Разделяющая поверхность

$$2ax_1 + 2bS^{-1}x_2 = C_1 - C_0$$