Машинное обучение: линейные модели final

Н. Поваров

* По материалам: Elements of Statistical Learning, Introduction to Statistical Learning, курсы ВШЭ, Википедия, личный опыт, ...

СПб,
2019

Что было до этого

- Эндогенность
- Автокорреляция
- Гетероскедастичность
- Мультиколлинеарность
- Свойства параметров линейной модели
- Closed form solution для линейных моделей

Как было до этого

- Поломка предпосылок
- Поломка свойств
- Как определять
- Как чинить
- С чем придётся смириться

Как будет сейчас

Разброд и шатание

Что мы пропустили

- Доверительные интервалы для параметров

Доверительные интервалы

- Асимптотические свойства

$$[\hat{\beta}_j - z_{cr}se(\hat{\beta}_j), \hat{\beta}_j + z_{cr}se(\hat{\beta}_j)]$$

- При нормальности ошибок

$$[\hat{\beta}_j - t_{cr}se(\hat{\beta}_j), \hat{\beta}_j + t_{cr}se(\hat{\beta}_j)]$$

Доверительные интервалы недостающее

-
- $\mathrm{se}(\hat{\beta}_j) = \sqrt{\widehat{Var}(\hat{\beta}_j|X)}$
- $\mathrm{Var}(\hat{\beta}_j|X) = \frac{\sigma^2}{RSS_j}$
- $\hat{\sigma}^2 = \frac{RSS}{n-m}$

Что мы пропустили

- ~~Доверительные интервалы для параметров~~
- Доверительные интервалы для прогноза

«Доверительные интервалы» для прогноза

- Доверительный интервал
- Предиктивный интервал

Доверительный интервал для прогноза

- Асимптотические свойства

$$[\hat{y}_i - z_{cr}se(\hat{y}_i), \hat{y}_i + z_{cr}se(\hat{y}_i)]$$

- При нормальности ошибок

$$[\hat{y}_i - t_{cr}se(\hat{y}_i), \hat{y}_i + t_{cr}se(\hat{y}_i)]$$

$$\mathrm{Var}(\hat{y}_i) = \mathrm{Var}(\hat{\beta}_0 + \hat{\beta}_1 x_i + \ \dots)$$

Предиктивный интервал для прогноза

- Асимптотические свойства

$$[\hat{y}_i - z_{cr}se(\hat{y}_i - \varepsilon_i), \hat{y}_i + z_{cr}se(\hat{y}_i - \varepsilon_i)]$$

- При нормальности ошибок

$$[\hat{y}_i - t_{cr}se(\hat{y}_i - \varepsilon_i), \hat{y}_i + t_{cr}se(\hat{y}_i - \varepsilon_i)]$$

$$\mathrm{Var}(\hat{y}_i - \varepsilon_i) = \mathrm{Var}\left(\hat{\beta}_0 + \hat{\beta}_1 x_i + \ \dots\right) + \mathrm{Var}(\varepsilon_i)$$

Что мы пропустили

- ~~Доверительные интервалы для параметров~~
- ~~Доверительные интервалы для прогноза~~
- Интерпретация значений параметров

Логарифмирование

- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
- $\ln(y_i) = \beta_0 + \beta_1 \ln(x_i) + \varepsilon_i$
- $\ln(y_i) = \beta_0 + \beta_1 x_i + \varepsilon_i$
- $y_i = \beta_0 + \beta_1 \ln(x_i) + \varepsilon_i$

Бинаризация факторов

- $\widehat{\text{Балл}}_i = \beta_0 + \beta_1 \text{ряд}_i + \beta_2 \text{рост}_i + \varepsilon_i$
- $\widehat{\text{Балл}}_i = \beta_0 + \beta_1 \text{ряд}_i + \beta_2 \text{рост}_i + \beta_3 \text{пол}_i + \varepsilon_i$
  - $\widehat{\text{Балл}}_i = (\beta_0 + \beta_3) + \beta_1 \text{ряд}_i + \beta_2 \text{рост}_i + \varepsilon_i$
  - $\widehat{\text{Балл}}_i = \beta_0 + \beta_1 \text{ряд}_i + \beta_2 \text{рост}_i + \varepsilon_i$
- $\widehat{\text{Балл}}_i = \beta_0 + \beta_1 \text{ряд}_i + \beta_2 \text{рост}_i + \beta_3 \text{пол}_i + \beta_4 \text{рост}_i \text{пол}_i + \varepsilon_i$
  - $\widehat{\text{Балл}}_i = (\beta_0 + \beta_3) + \beta_1 \text{ряд}_i + (\beta_2 + \beta_4) \text{рост}_i + \varepsilon_i$
  - $\widehat{\text{Балл}}_i = \beta_0 + \beta_1 \text{ряд}_i + \beta_2 \text{рост}_i + \varepsilon_i$

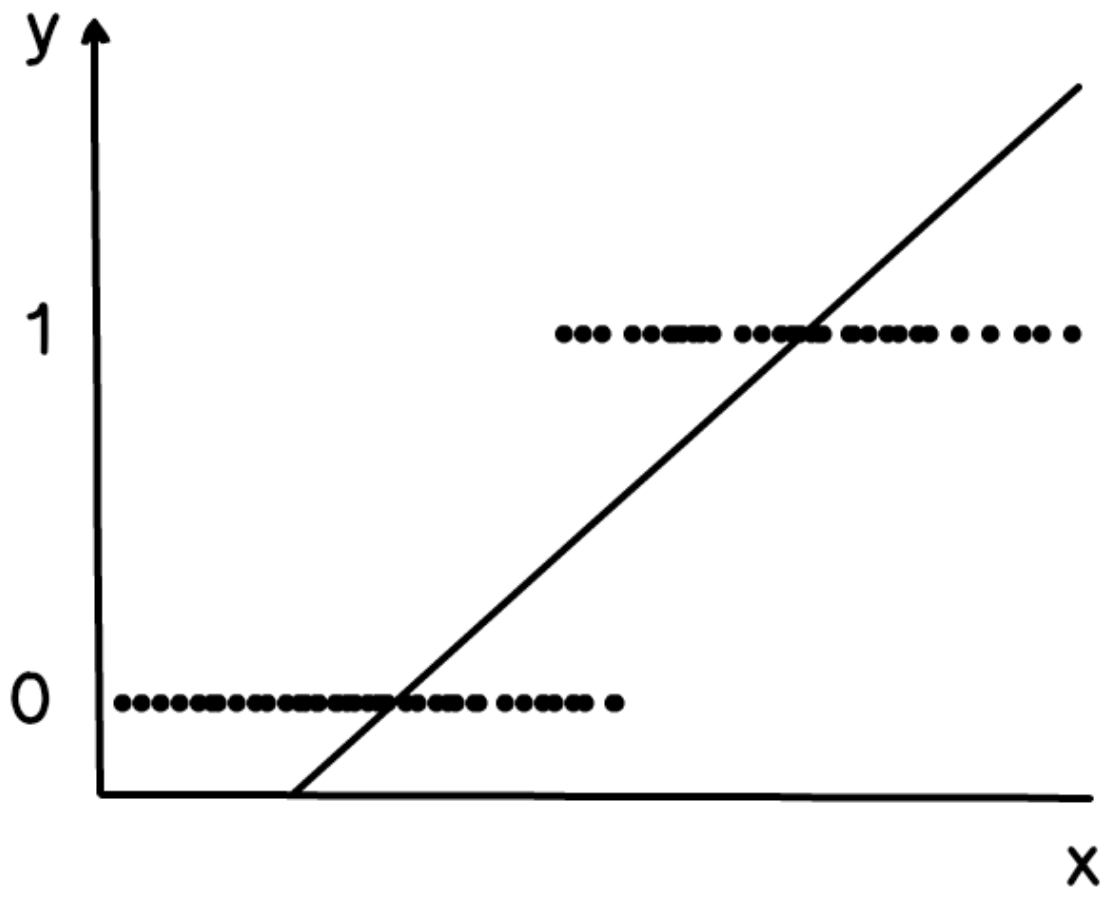
Что мы пропустили

- ~~Доверительные интервалы для параметров~~
- ~~Доверительные интервалы для прогнозов~~
- ~~Интерпретация значений параметров~~
- Классификация

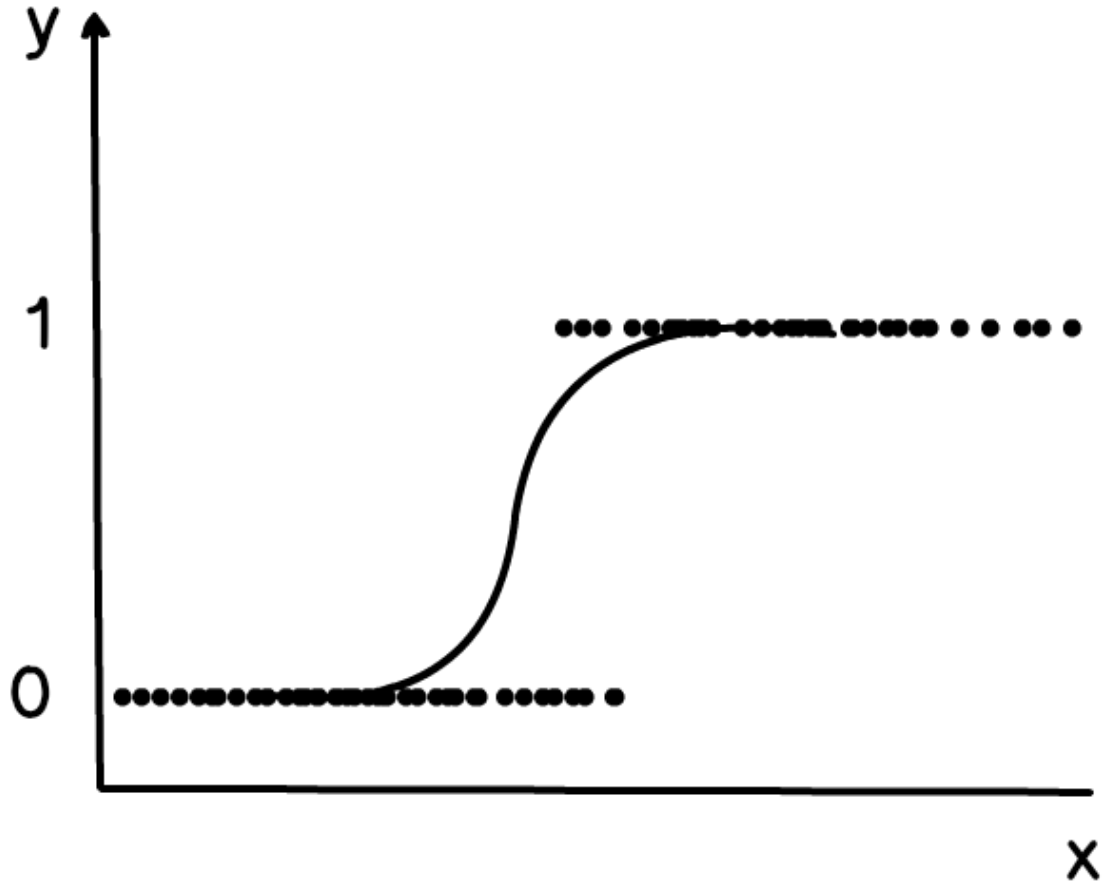
Задача классификации

- Иногда перед нами стоит задача классификации
- Регрессия может быть классификацией
- Хотим вероятности

Пара картинок



Пара картинок



Логистическая регрессия

•

$$h(z) = \frac{1}{1 + e^{-z}} = \frac{e^{z}}{1 + e^{z}}$$

$$z = \beta_0 + \beta_1 x_i + \cdots + \varepsilon_i$$

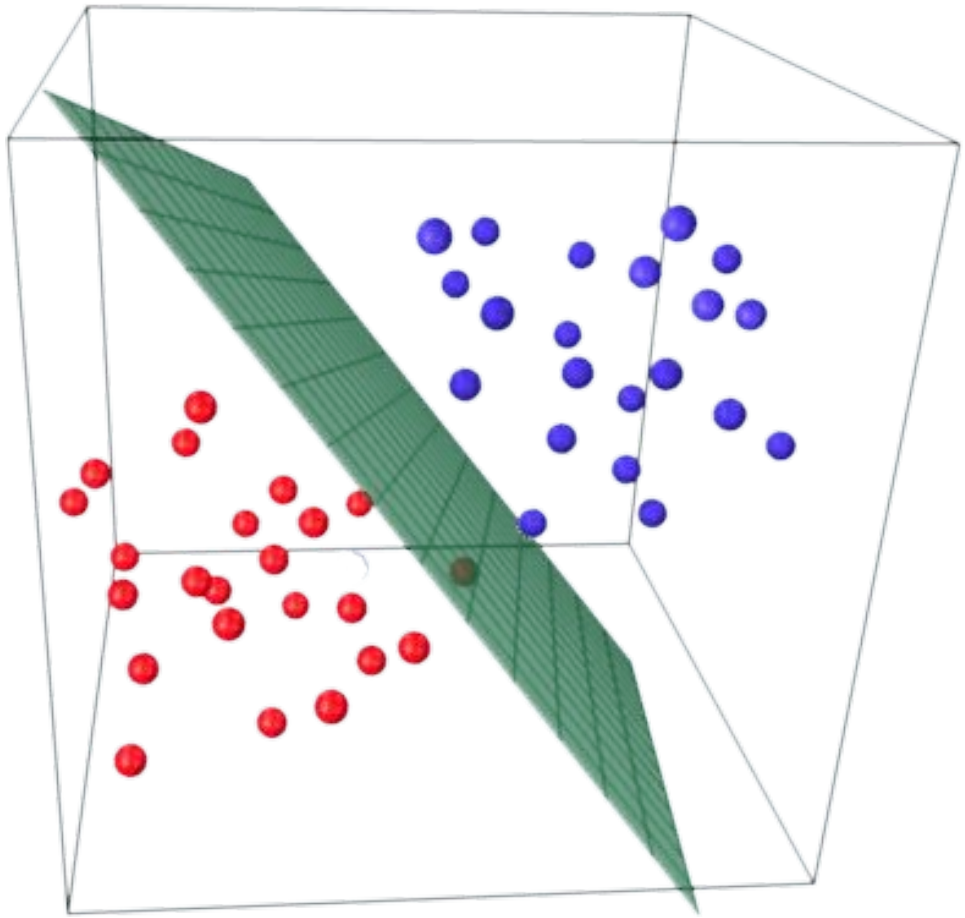
$$\ln \frac{p(y_i = 1)}{p(y_i = 0)} = \beta_0 + \beta_1 x_i + \cdots + \varepsilon_i$$

Как найти параметры

•

$$\beta_j := \beta_j + \alpha(y_i - h_\beta(\overrightarrow{x_i}))x_j$$

Ещё один взгляд на классификацию

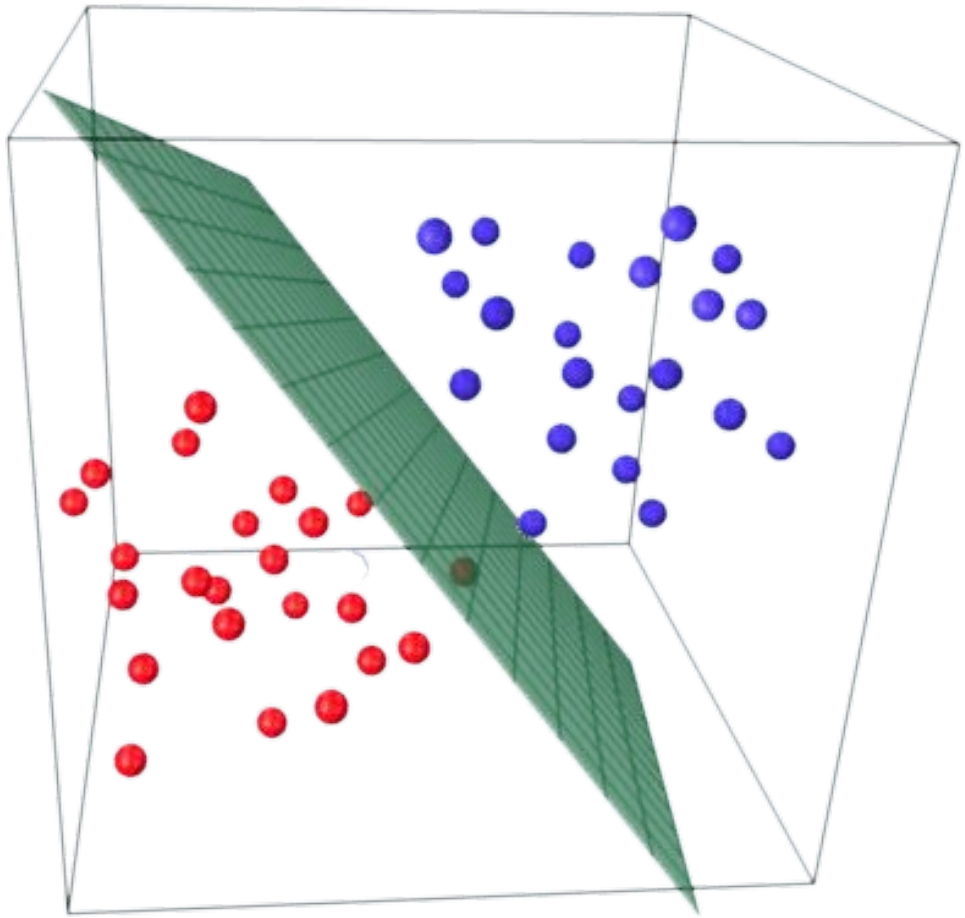


Попишем формулы

- $$Q = \frac{1}{n}\sum_{i=1}^{n}\left[\mathrm{F}_{\beta}(\overrightarrow{\mathrm{x}_{\mathrm{i}}}) \neq \mathrm{y}_{\mathrm{i}}\right]$$

$$Q = \frac{1}{n}\sum_{i=1}^{n}\left[\mathrm{y}_{\mathrm{i}}\left(\vec{\beta}, \overrightarrow{\mathrm{x}_{\mathrm{i}}}\right) < 0\right]$$

Ещё один взгляд на классификацию



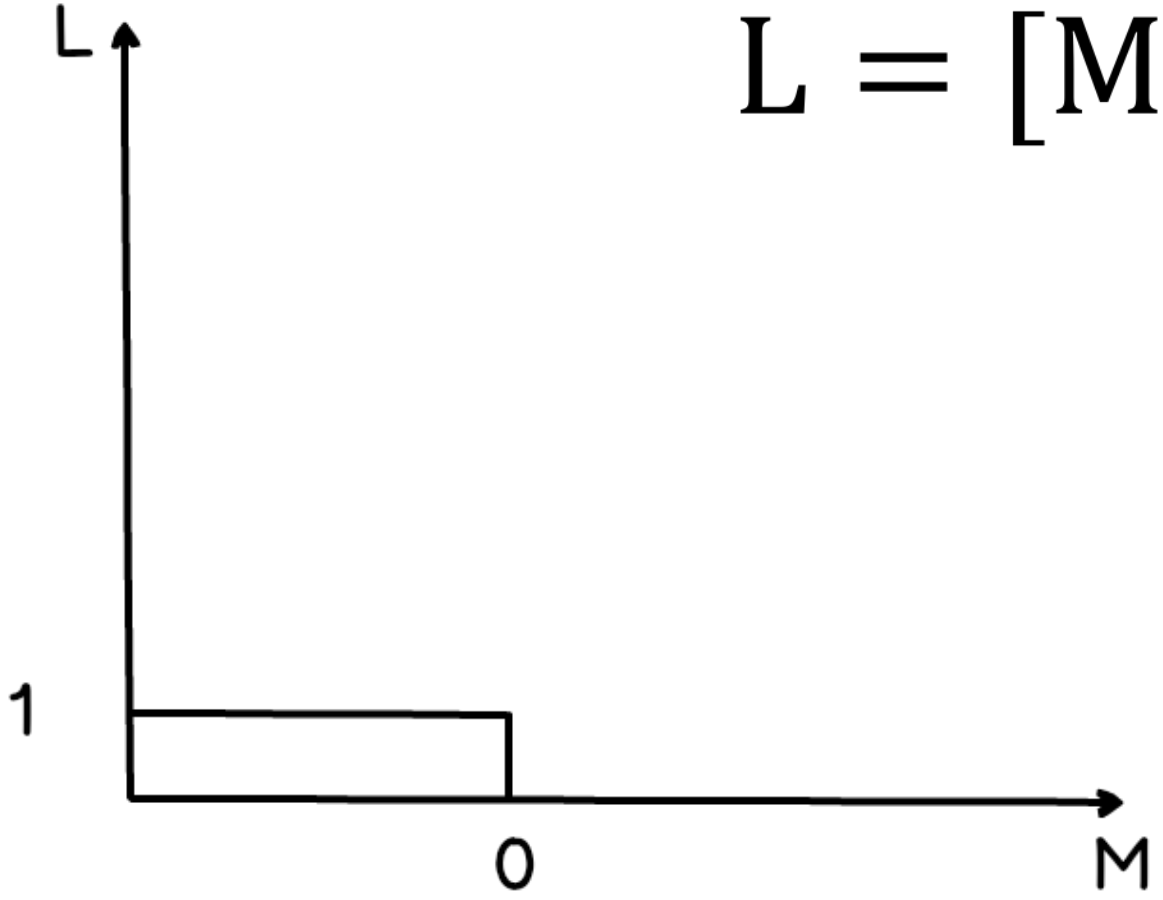
Добавим обозначений

- $$Q = \frac{1}{n}\sum_{i=1}^{n}\left[\mathrm{y_i}\left(\vec{\beta}, \overrightarrow{\mathrm{x_i}}\right) < 0\right]$$

$$M = \mathrm{y_i}\left(\vec{\beta}, \overrightarrow{\mathrm{x_i}}\right)$$

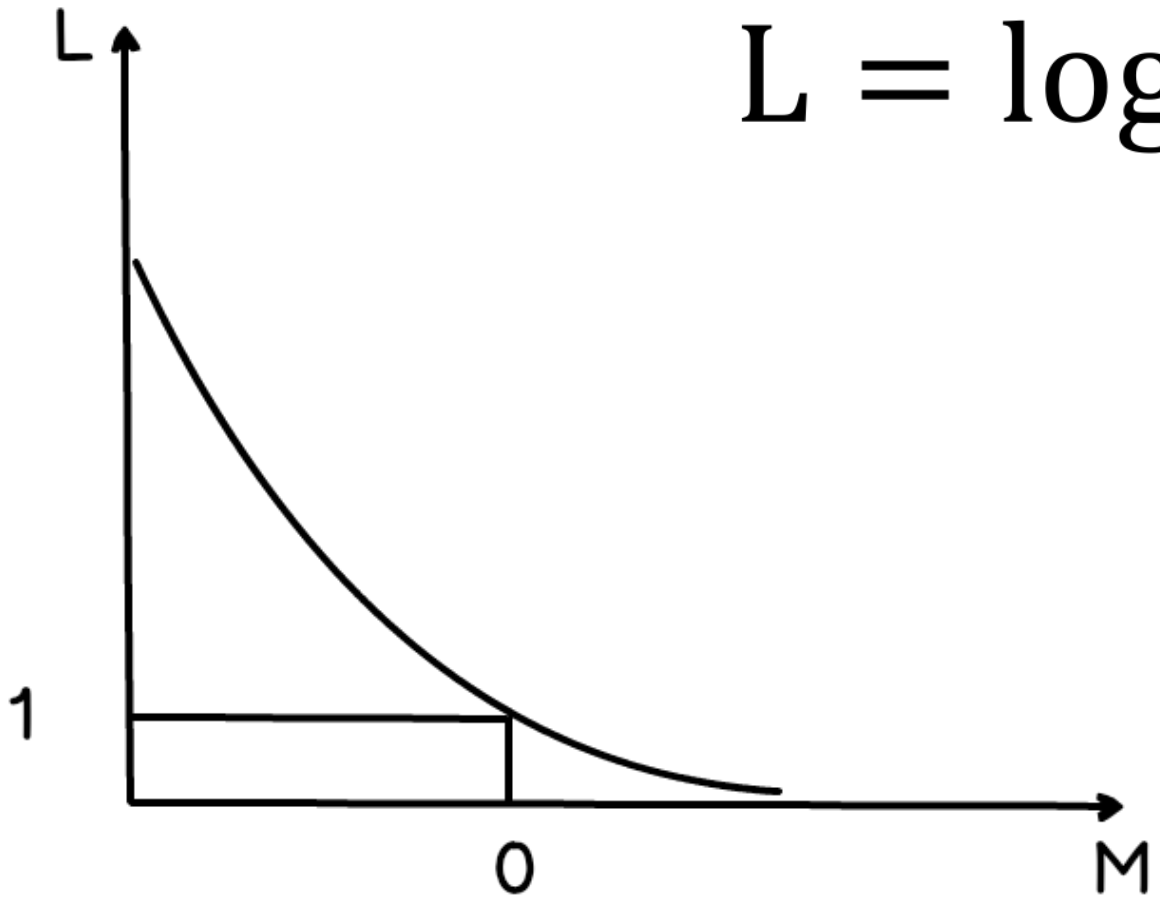
$$Q = \frac{1}{n}\sum_{i=1}^{n} L$$

Функция потерь



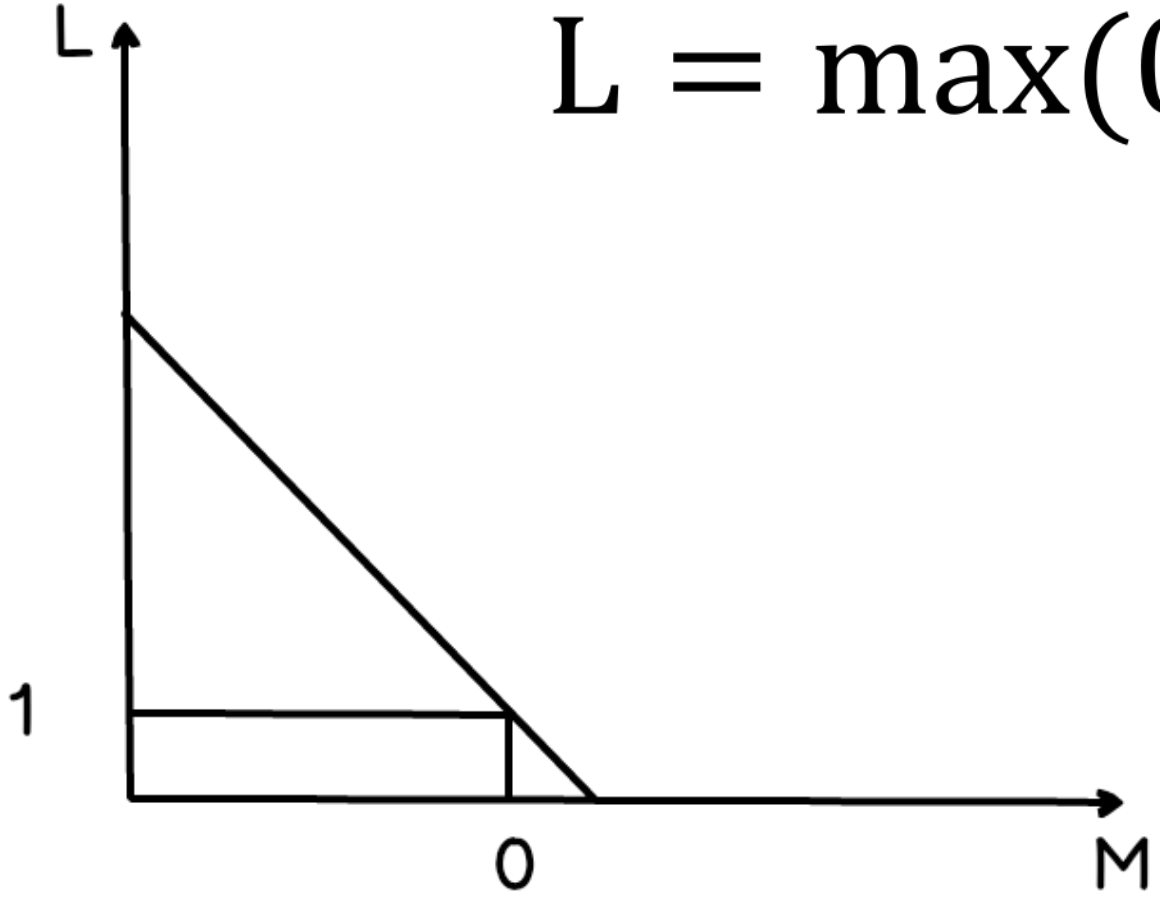
$$L = [M < 0]$$

Логистическая функция



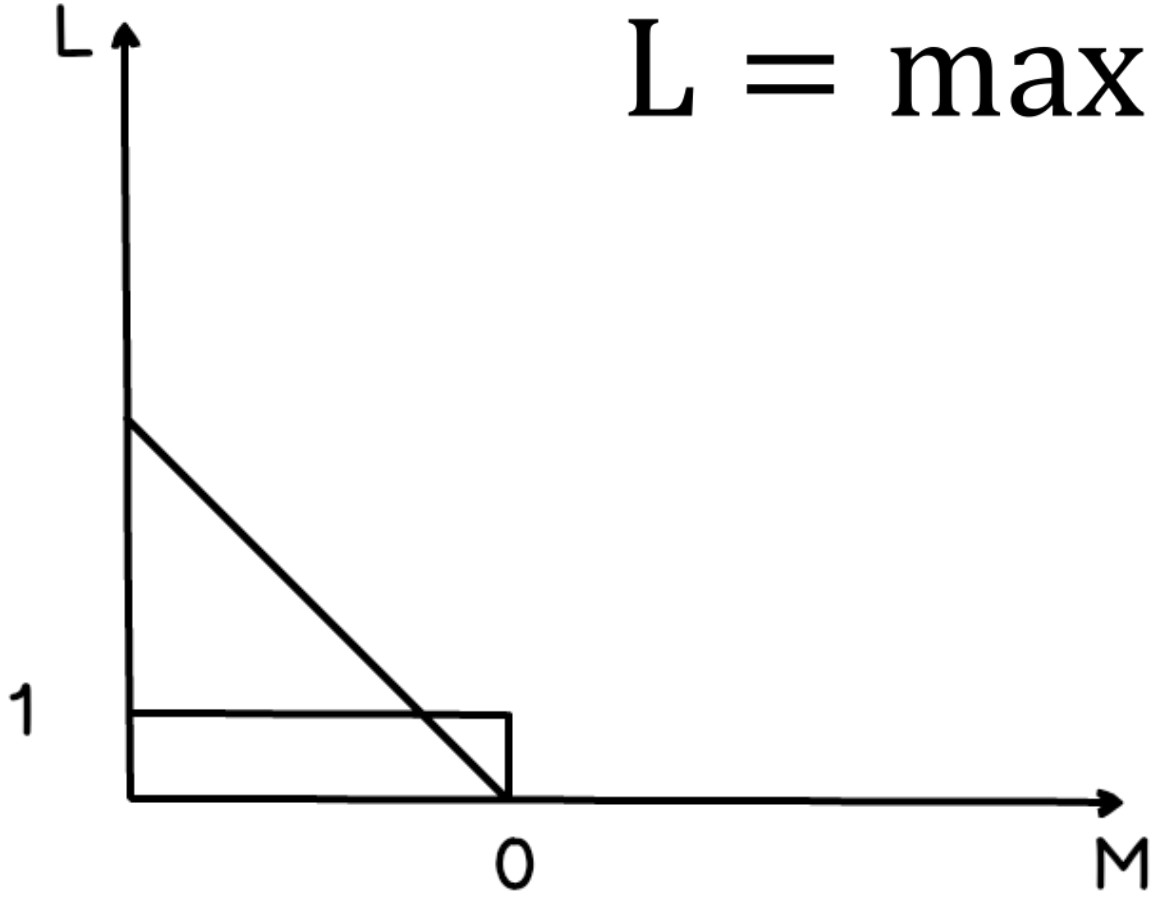
$$L = \log(1 + e^{-M})$$

Hinge Loss



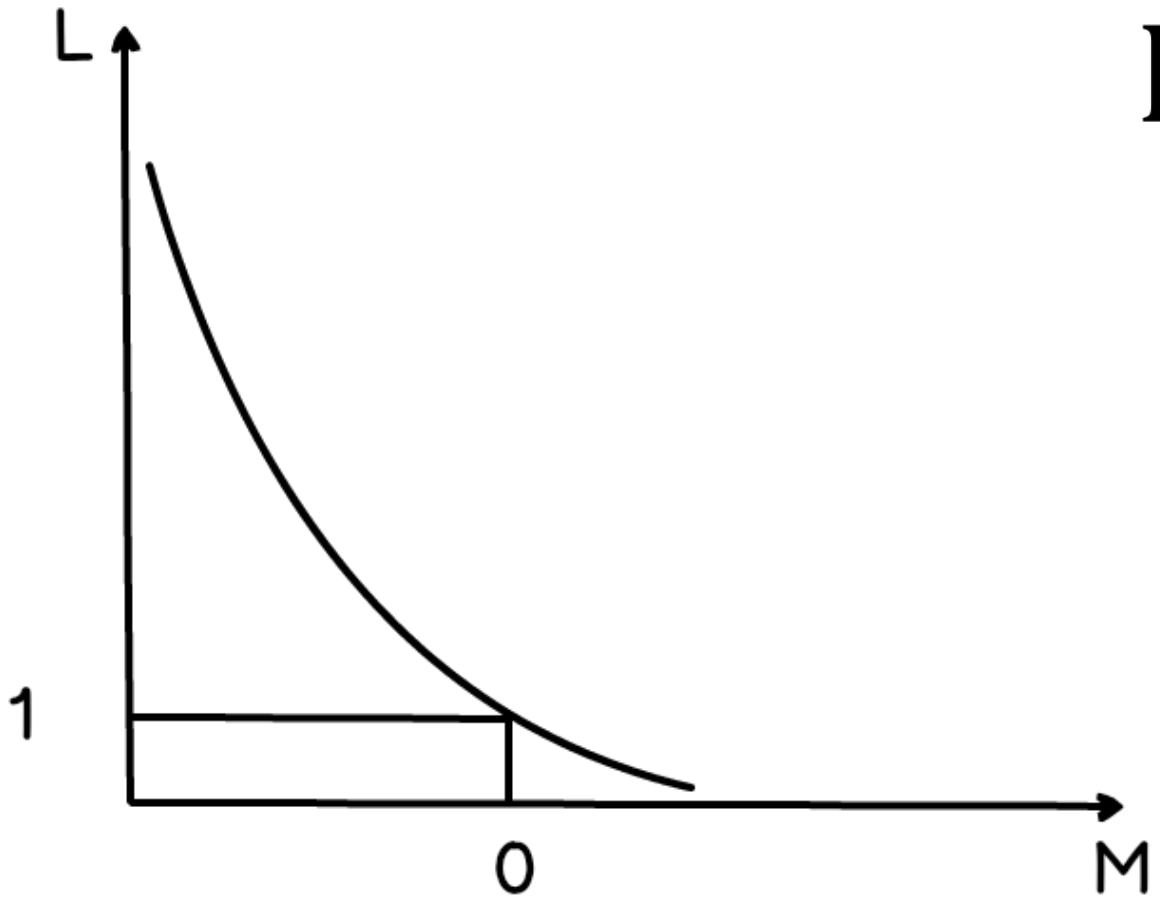
$$L = \max(0,1 - M)$$

В перцептронах



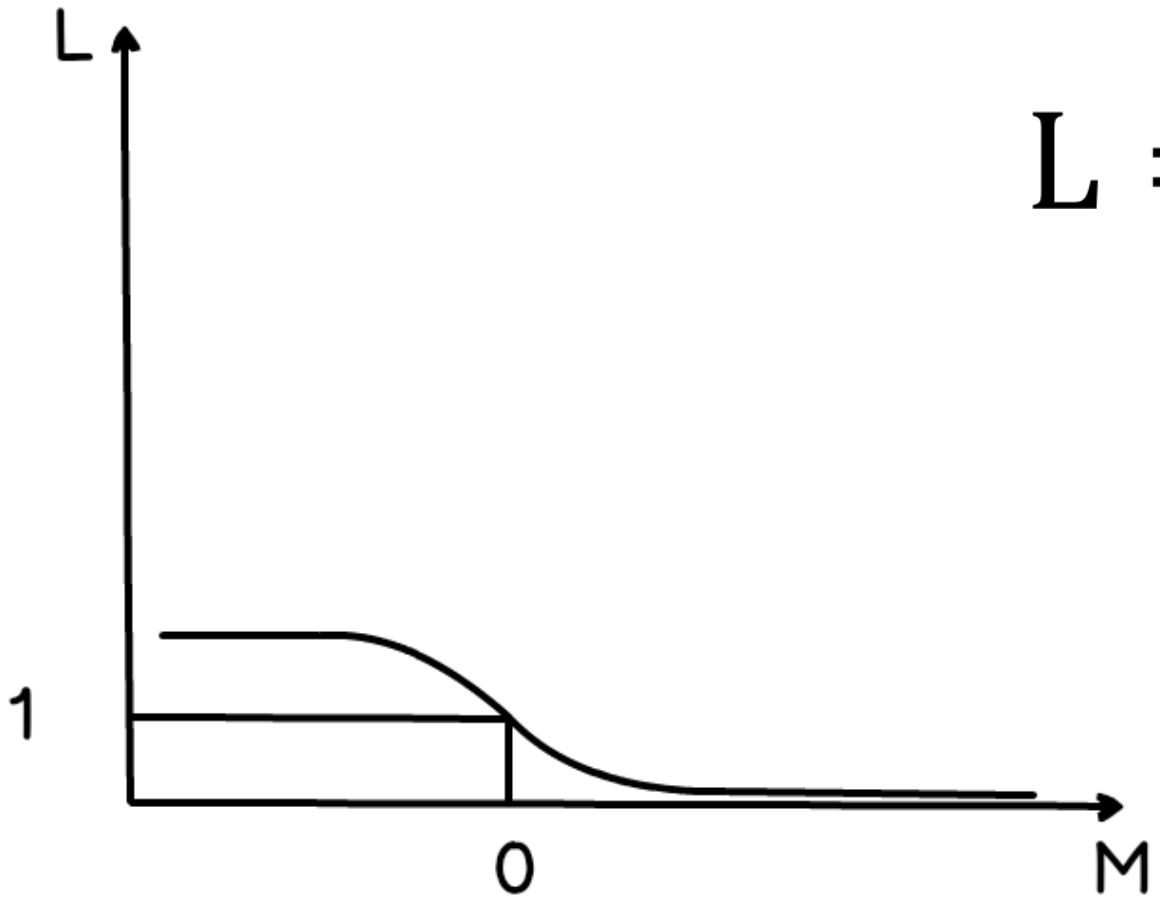
$$L = \max(0, -M)$$

Экспоненциальная



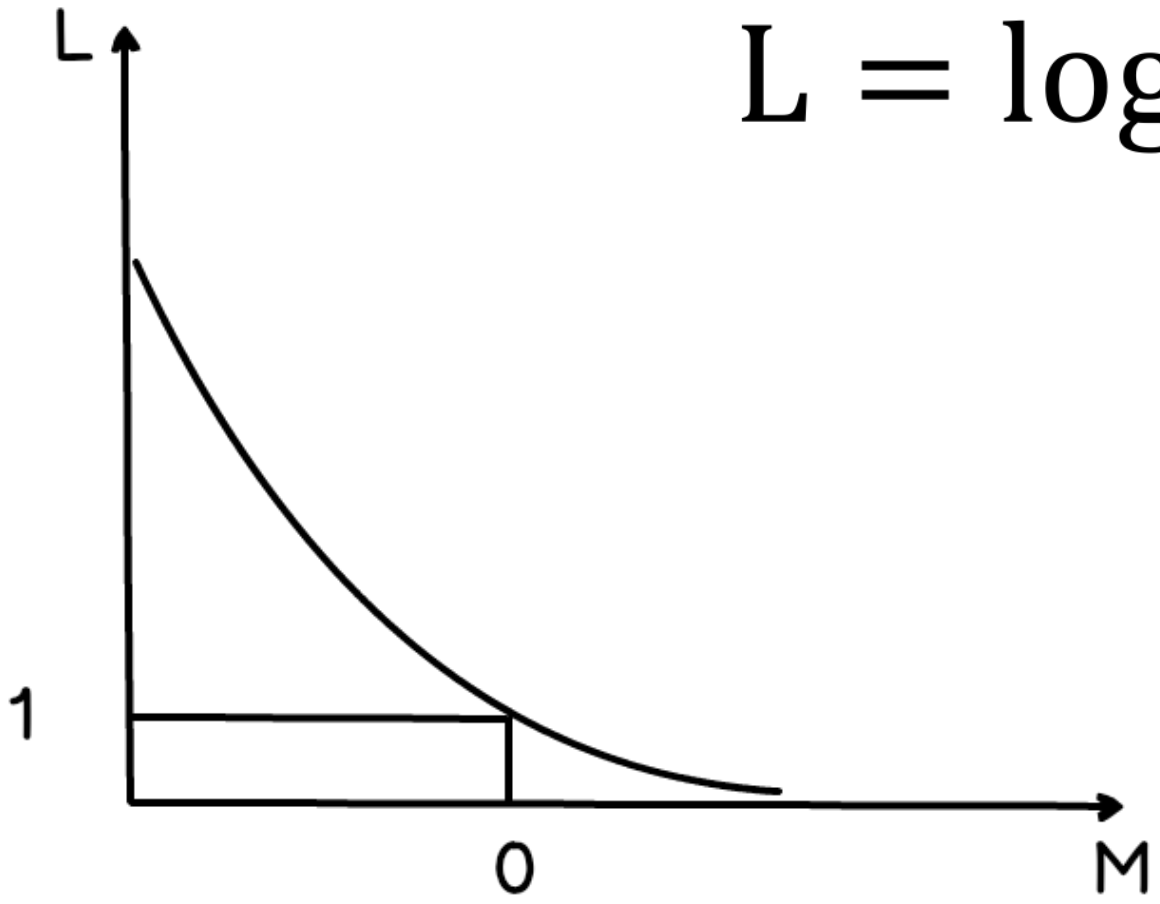
$$L = e^{-M}$$

Сигмоидная



$$L = \frac{2}{1 + e^{M}}$$

Логистическая функция



$$L = \log(1 + e^{-M})$$

Hinge Loss в явном виде

•

$$\min_{\vec{\beta}} \sum_{i=1}^{n} \max(0, 1 - y_i(\vec{\beta}, \vec{x_i}))$$

Hinge Loss + L2 регуляризация

•

$$\min_{\vec{\beta}} \sum_{i=1}^{n} \max(0, 1 - y_i(\vec{\beta}, \vec{x_i})) + \lambda \sum_{j=1}^{m} \hat{\beta}_j^2$$

Hinge Loss + L2 регуляризация

•

$$\min_{\vec{\beta}} C \sum_{i=1}^{n} \max(0,1 - y_i(\vec{\beta}, \vec{x_i})) + \frac{1}{2} \sum_{j=1}^{m} \hat{\beta}_j^2$$

~~Hinge Loss + L2 регуляризация~~

•

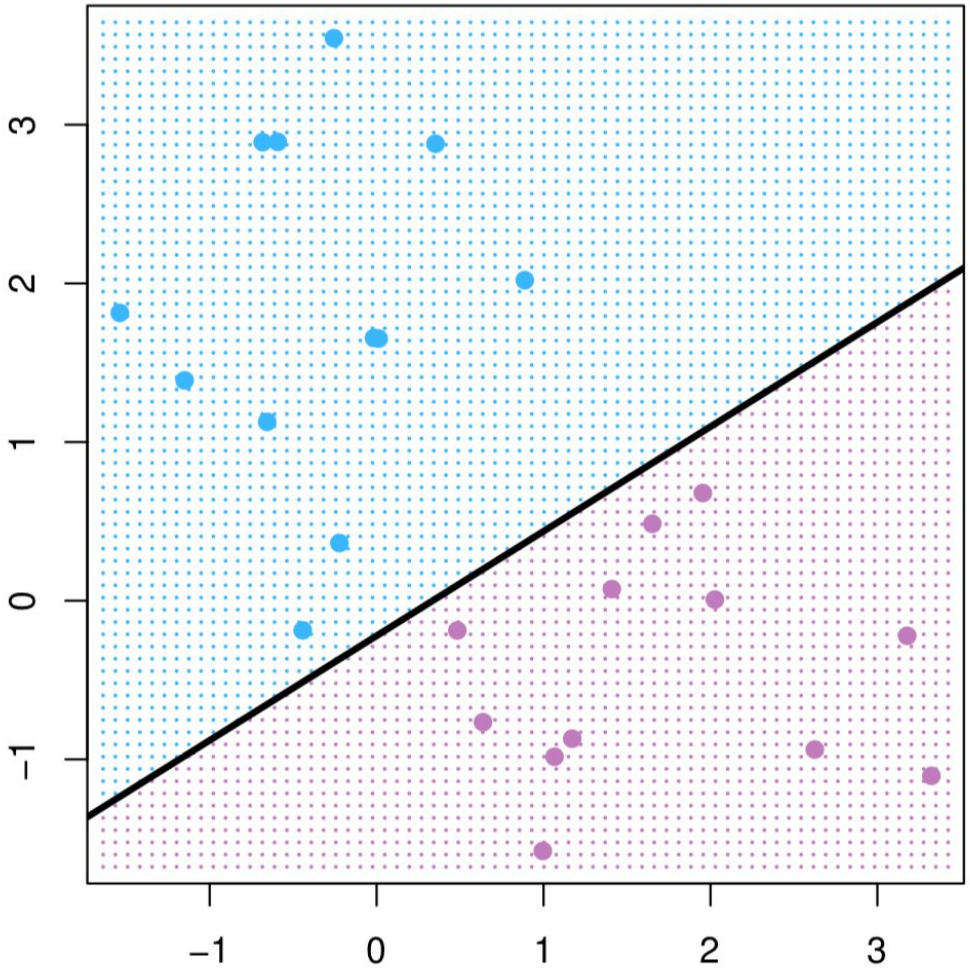
$$\min_{\vec{\beta}} C \sum_{i=1}^{n} \max(0,1 - y_i(\vec{\beta}, \overrightarrow{x_i})) + \frac{1}{2} \sum_{j=1}^{m} \hat{\beta}_j^2$$

SVM

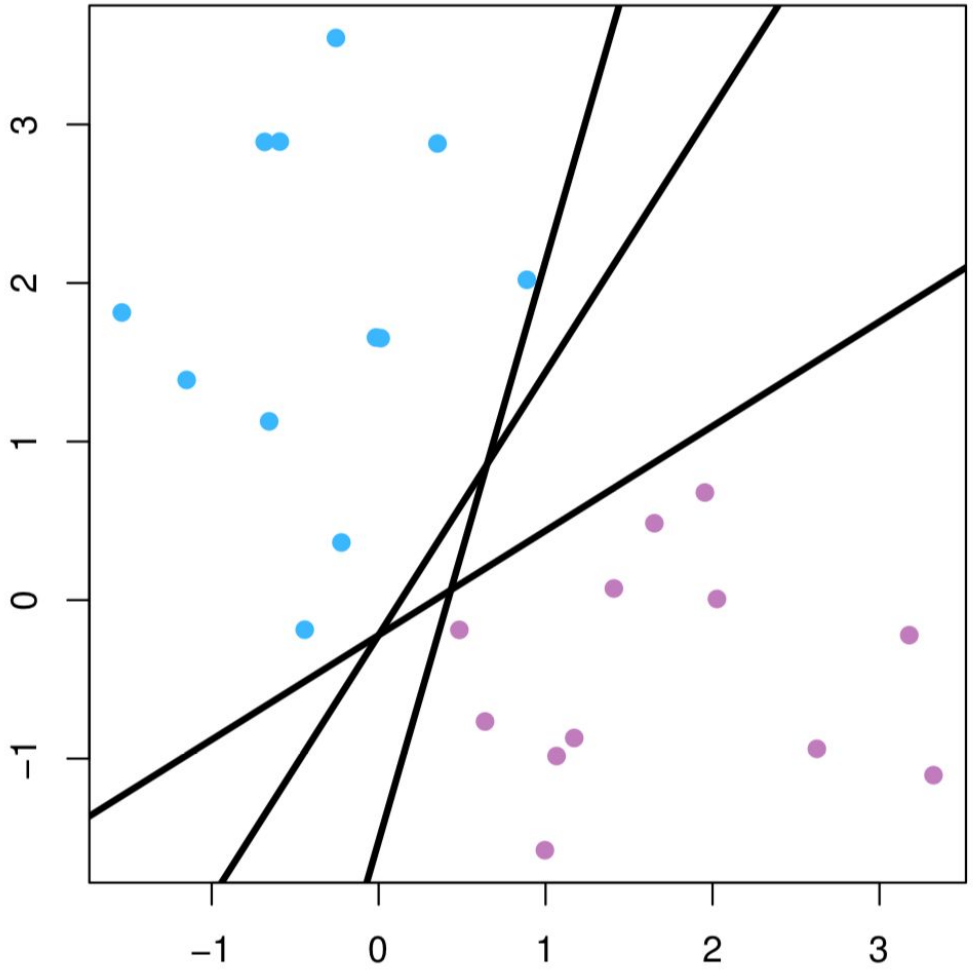
•

$$\min_{\vec{\beta}} C \sum_{i=1}^{n} \max(0,1 - y_i(\vec{\beta}, \vec{x_i})) + \frac{1}{2} \sum_{j=1}^{m} \hat{\beta}_j^2$$

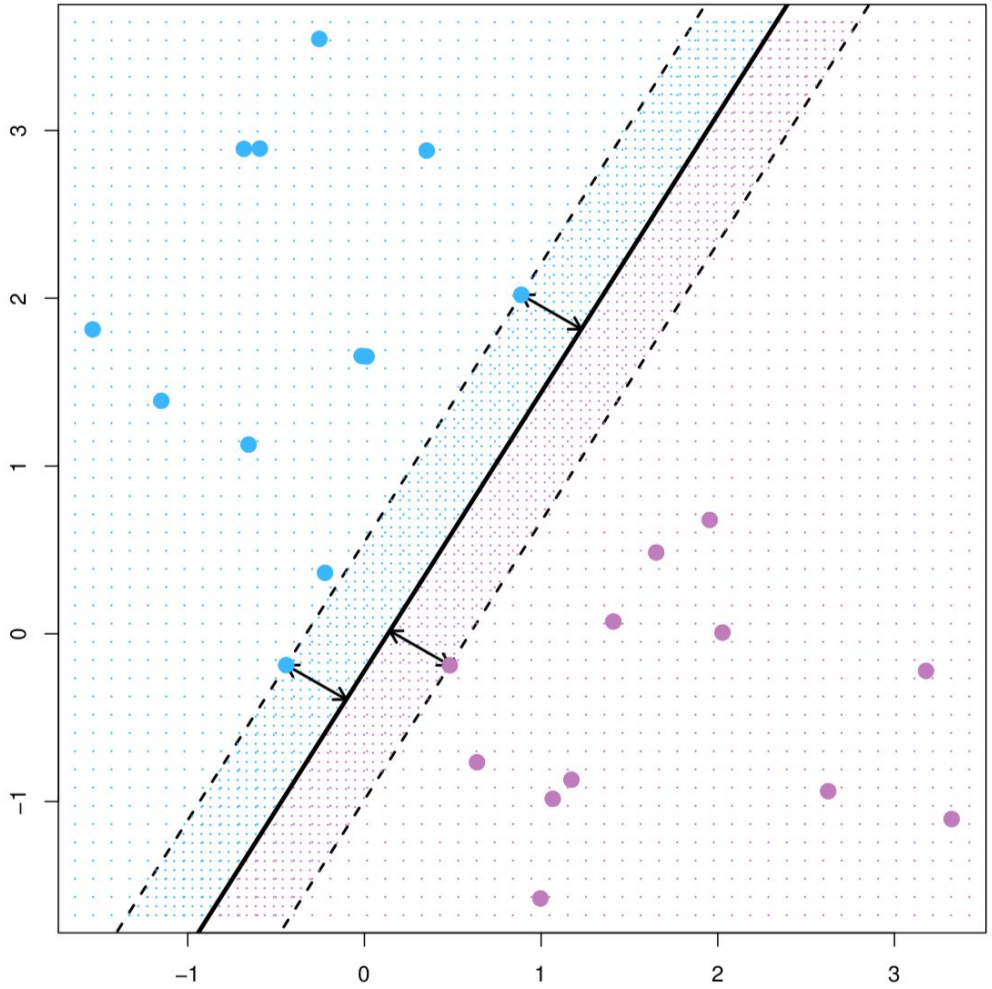
SVM исторически



SVM исторически



SVM исторически



Оценка качества классификатора

y

		P	N
$\hat{y}$	P	TP	FP
	N	FN	TN

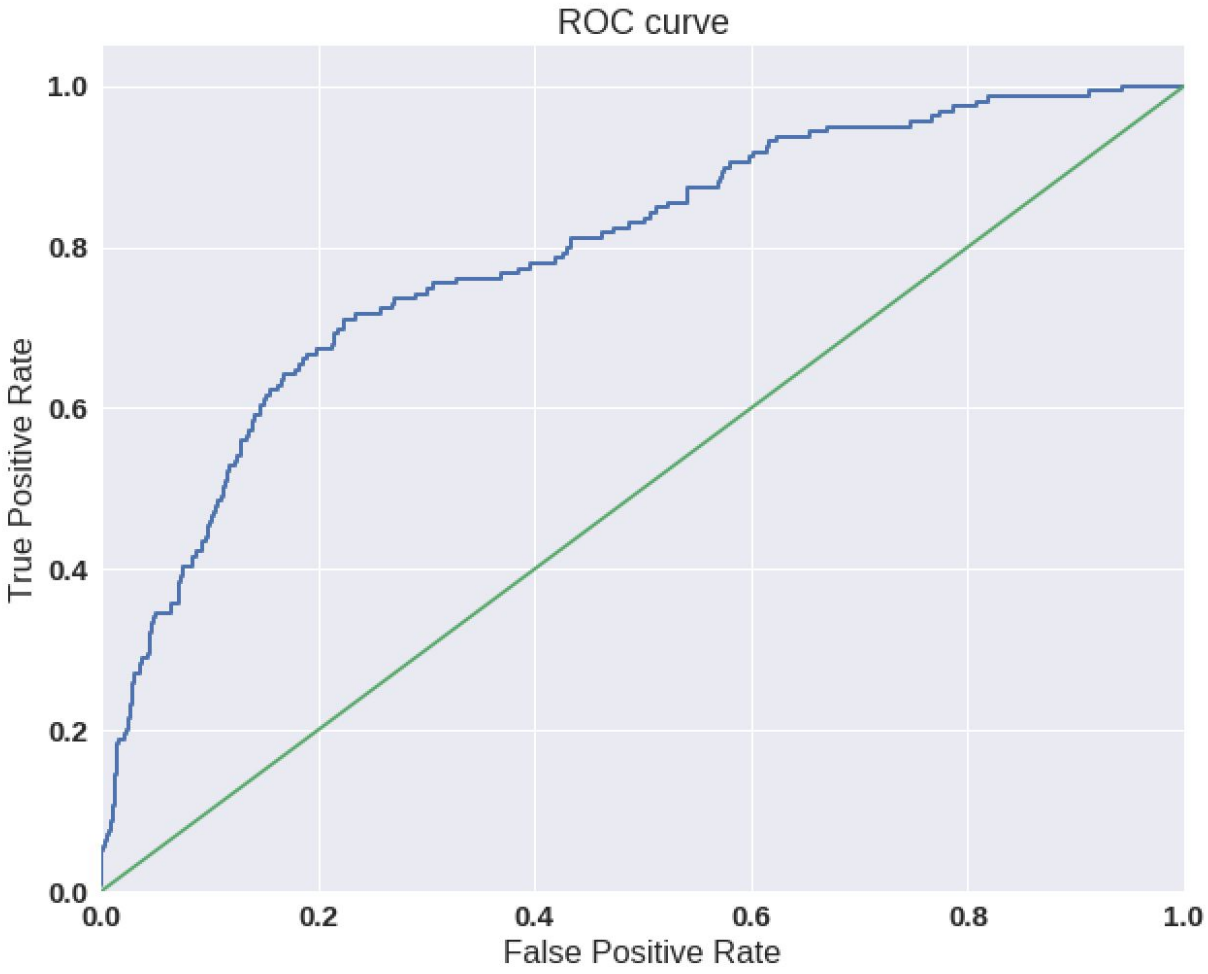
Метрики классификатора

-
- $\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$
- $\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$
- $\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FN+TN+FN}}$
- $\text{F}_1 = \frac{2*\text{Pr}*\text{Re}}{\text{Pr} + \text{Re}}$

Хорошие метрики классификатора

- AUC
- LL

AUC



$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

LogLoss

•

$$LL = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)$$

Что мы вспомнили

- Доверительные интервалы для параметров
- Доверительные интервалы для прогнозов
- Интерпретация значений параметров
- Логистическая регрессия
- SVM
- Метрики классификации

Что не рассказано

- SVM регрессия
- SVM kernel trick
- Свойства параметров логистической регрессии
- Мультиклассификация
- Временные ряды
- Медианная и квантильная регрессия
- Compressed Sensing
- Байесовский подход
- …

Вопросы?