Ансамбли

Корлякова Мария.

2021

Регрессионная модель связи X и D

- $D = f(X) + \varepsilon$
- ε - ожидаемая ошибка
  - нормально распределена
  - $M(\varepsilon)=0$
- $f(X)$ – регрессионная модель
- она есть объективно

Свойства модели

1. *Среднее значение ожидаемой ошибки ε для любой реализации X $E[\varepsilon|x] = 0$, тогда*
2. $f(x) = E[D|x]$
3. *Ошибка ε не коррелирует с функцией регрессии f(X):* $E[\varepsilon f(X)]=0$

E[ ] - математическое ожидание

Определение w

- Минимизация функции стоимости:

  $F(x_i,w)$ – модель (мы ее строим)

  Критерий:

  $Er(w) = ½ \Sigma (d_i – F(x_i,w))^2$,

  $Er(w) = ½ E_T[(f(x) – F(x,T))^2]$

  W зависят от выборки T={(X,d)}

  Et[ ] - среденее выборочное

Мера прогнозирования

$$L_{av}(f(x), F(x,T)) = E_T[(f(x) - F(x,T))^2]$$

f(x) = E[D|x] – f(x) мат. ожидание D|x

$$L_{av}(f(x), F(x,T)) = E_m[(E[D|X=x] - F(x,T))^2]$$

Ошибка оценивания регрессионной функции f(X) аппроксимационной F(x,T)

$$(E[D|X=x] - F(x,T)) =$$

$$(E[D|X=x] - E_T[F(x, T)]) + (E_T[F(x, T)] - F(x,T))$$

Тогда

$$L_{av}(f(x),F(x,T)) = E_T[(E[D|X=x] - F(x,T))^2] = B^2(w) + V(w) +$$

$$+ E_T[(E[D|X=x] - f(x))^2]$$

* пренебрегаем бесконечно малыми и получим простое выражение

$B(w) = E_T\,[(E[D|X=x] - F(x,\ T))\ ]$ – смещение среднего для $F(x,\ T)$ *относительно* $f(x) \Rightarrow$ ***ошибка аппроксимации***

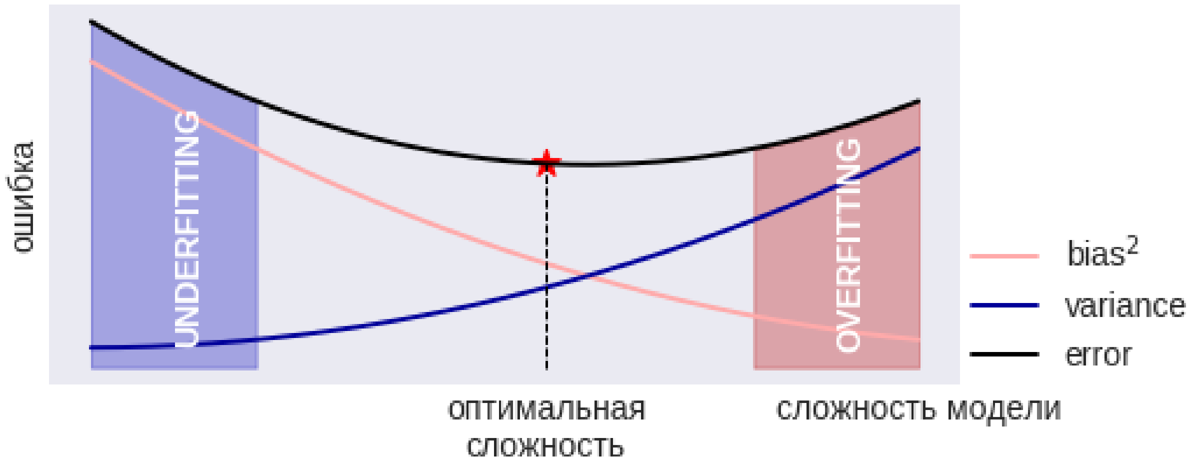
$V(w) = E_T\,[\ (E_T[F(x,\ T)] - F(x,T))^2\ ]$ – ***дисперсия*** $F(x,T)$ *на всем* T.

Дилемма дисперсии-смещения

Дилемма дисперсии-смещения

Дилемма дисперсии-смещения

- Одновременно уменьшить смещение и дисперсию можно только для бесконечно большой выборки



Как решить дилему:

- Композиции алгоритмов
  - Алгебраический подход к построению корректных алгоритмов
    - Области компетентности
    - Багинг - bagging
    - Бустинг - boosting

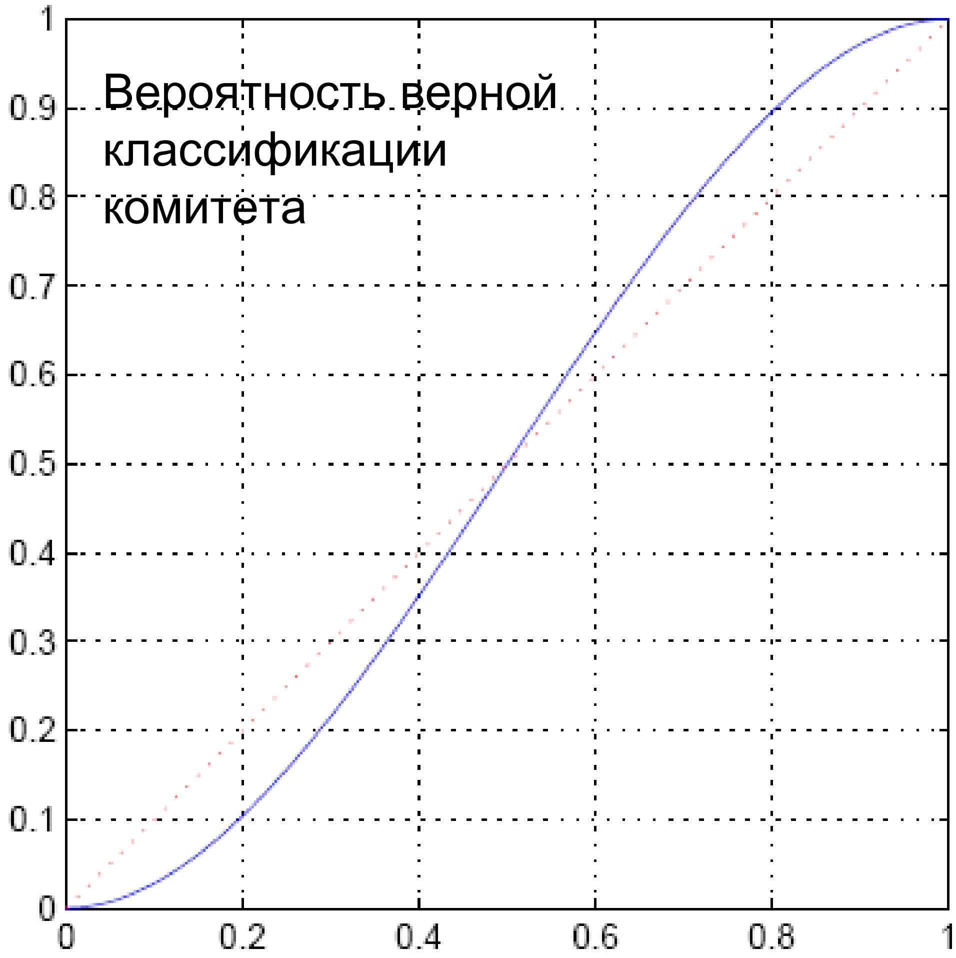
Центральная предельная теорема (теор.вер.)

- Последовательности частичных средних, вычисленных по наборам из n независимых случайных величин, даже имеющих большую дисперсию σ, стремятся к нормальному распределению с дисперсией $D=D/\sqrt{n}$.
- Тем самым, если использовать в качестве результата среднее от значений прогнозов отдельных моделей, то неопределенность такого результата окажется ниже неопределенности отдельной модели.

- Пусть есть три алгоритма A1, A2, A3
- Ai, решает определенную задачу бинарной классификации с вероятностью успеха **p**, независимо от остальных.
- Тогда при классификации примера X возможны 8 исходов:
  - все классификаторы выдали верный ответ, $\mathbf{p^3}$
  - два из трех не ошиблись (три варианта), $\mathbf{3p^2(1 - p)}$**,**
  - не ошибся лишь один (еще три варианта), $\mathbf{3p(1 - p)^2}$
  - ошиблись все три алгоритма одновременно, $\mathbf{(1 - p)^3}$.

- *комитет большинства*,
- вероятность благоприятного исхода (1 и 2 вариант)
- $q = p^3 + 3p^2(1 - p) = 3p^2 - 2p^3$.

Вероятность верной классификации комитетом



Точность одного классификатора

Усреднение по ансамблю

- Стратегия обучения
  - Уменьшение общей ошибки за счет варьирования начальных состояний.
  - Эксперты обучаются с избытком
  - Дисперсия уменьшается за счет усреднения

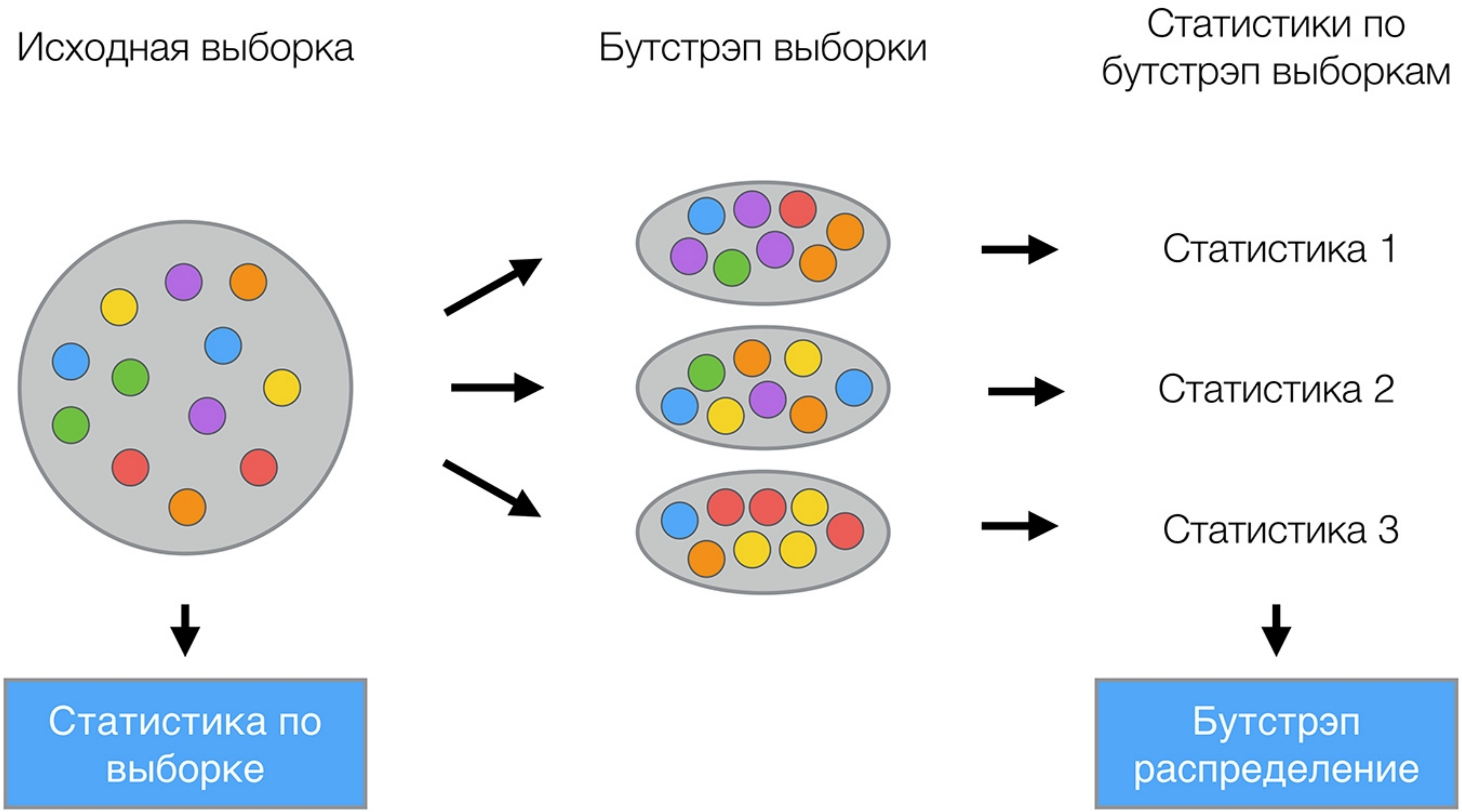
точность ансамбля моделей

- можно улучшить, если:

1) повысить точность каждой отдельной модели и, одновременно,

2) обеспечить статистическую независимость ошибок разных членов ансамбля.

Бэгинг



Исходная выборка
Бутстрэп выборки
Статистики по бутстрэп выборкам
Статистика 1
Статистика 2
Статистика 3
Статистика по выборке
Бутстрэп распределение

Out-Of-Bag-Error