	Задача	Модель f	Параметры $\mathbf{w}$	Гиперпараметры $\mathbf{A}, \mathbf{B}$	Признаки $\mathcal{A}$	Объекты $\mathcal{B}$
1	Оценка параметров	задана	найти	–	заданы	заданы
2	Выбор оптимальной модели	найти	заданы	–	заданы	заданы
3	Оценка ковариационных матриц	задана	–	найти	заданы	заданы
4	Выбор объектов и признаков	задана	–	заданы	найти подмножество	найти подмножество
5	Выбор правдоподобной модели	найти	–	заданы	заданы	заданы
6	Выбор смеси моделей	задана	–	заданы	заданы	найти покрытие
7	Нахождение инвариантов	задана	–	заданы	заданы	найти разбиение
8	Оценка мощности выборки	задана	заданы	заданы	заданы	оценить число
9	Проверка гипотезы порождения данных	задана	заданы	–	–	заданы

Таблица 3. Сводная таблица задач, решаемых при восстановлении регрессии.

1.4.1. Линейные модели

Нахождение параметров $\mathbf{w}$ линейной модели (3) при предположении о нормальном распределении (15) зависимой переменной $\mathbf{y}$ заключается в минимизации евклидовой нормы вектора регрессионных остатков

$$S(\mathbf{w}) = \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2 = \|\boldsymbol{\varepsilon}\|^2. \tag{36}$$

Предполагается выполнение следующих условий:

(37)

1) независимые переменные $\mathbf{x}$ не являются случайными величинами,

2) математическое ожидание $\mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$,

3) дисперсия $\mathsf{D}(\boldsymbol{\varepsilon}) = \sigma_{\boldsymbol{\varepsilon}}^2\mathbf{I}$ (условие гомоскедастичности),

4) при $i \neq k$ математическое ожидание $\mathsf{E}(\varepsilon_i, \varepsilon_k) = 0$,

5) $\mathrm{rank}(\mathbf{X}) = n \leqslant m$.

Эти условия называются условиями Гаусса-Маркова [19]. При этом оценки параметров модели (3) являются состоятельными и несмещенными. Оценки являются также эффективными, если $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{\varepsilon}}^2\mathbf{I})$.

Требуется минимизировать евклидово расстояние от вектора $\mathbf{y}$ до вектора $\mathbf{Xw}$. Этот вектор лежит в пространстве столбцов матрицы $\mathbf{X}$, так как $\mathbf{Xw}$ — это линейная комбинация столбцов этой матрицы с коэффициентами $w_1, \ldots, w_n$. Задача оценки $\mathbf{w}$ эквивалентна задаче нахождения точки $\mathbf{p} = \mathbf{X}\mathbf{w}$, ближайшей к $\mathbf{y}$ и находящейся в пространстве столбцов матрицы $\mathbf{X}$. Следовательно, вектор $\mathbf{p}$ должен быть проекцией $\mathbf{y}$ на пространство столбцов, вектор регрессионных остатков $\mathbf{X}\mathbf{w} - \mathbf{y}$ должен быть ортогонален этому пространству. Рассмотрим произвольный вектор $\mathbf{X}\mathbf{v}$, ортогональный вектору регрессионных остатков $\mathbf{X}\mathbf{w} - \mathbf{y}$:

$$(\mathbf{X}\mathbf{v})^{\mathsf{T}}(\mathbf{X}\mathbf{w} - \mathbf{y}) = \mathbf{v}^{\mathsf{T}}(\mathbf{X}^{\mathsf{T}}\mathbf{X}\mathbf{w} - \mathbf{X}^{\mathsf{T}}\mathbf{y}) = 0.$$

Так как это равенство должно быть справедливо для произвольного вектора $\mathbf{v}$, то $\mathbf{X}^{\mathsf{T}}\mathbf{X}\mathbf{w} - \mathbf{X}^{\mathsf{T}}\mathbf{y} = 0$, см. рис. 6. Если столбцы матрицы $\mathbf{X}$ линейно независимы, то матрица $\mathbf{X}^{\mathsf{T}}\mathbf{X}$ обратима и уравнение имеет единственное решение относительно параметров

$$\mathbf{w} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{y}. \tag{38}$$

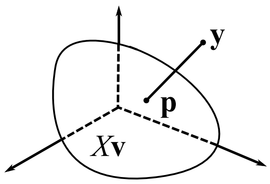


Рис. 6. Проекция вектора $\mathbf{y}$ на пространство столбцов матрицы $\mathbf{X}$.

Проекция вектора $\mathbf{y}$ на пространство столбцов матрицы $\mathbf{X}$ имеет вид

$$\mathbf{p} = \mathbf{X}\mathbf{w} = \mathbf{X}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{y} = \mathbf{P}\mathbf{y}.$$

Матрица $\mathbf{P} = \mathbf{X}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}$ называется матрицей проектирования. Она она идемпотентна, $\mathbf{P}^2 = \mathbf{P}$, и симметрична, $\mathbf{P}^{\mathsf{T}} = \mathbf{P}$.

Используемые здесь методы нахождения оптимальных параметров моделей предполагают непрерывную дифферренцируемость функции $S(\mathbf{w})$ в области $\mathcal{W} \ni \mathbf{w}$. Согласно (36),

$$S(\mathbf{w}) = (\mathbf{X}\mathbf{w} - \mathbf{y})^{\mathsf{T}}(\mathbf{X}\mathbf{w} - \mathbf{y}) = \mathbf{y}^{\mathsf{T}}\mathbf{y} - 2\mathbf{y}^{\mathsf{T}}\mathbf{X}\mathbf{w} + \mathbf{w}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{X}\mathbf{w}.$$

Для того, чтобы найти минимум этой функции, требуется приравнять её градиент к нулю:

$$\frac{\partial S}{\partial \mathbf{w}} = -2\mathbf{X}^{\mathsf{T}}\mathbf{y} + 2\mathbf{X}^{\mathsf{T}}\mathbf{X}\mathbf{w} = 0.$$

Решение этого уравнения совпадает с решением (38).

Validation

Consider partition $\mathcal{I} = \mathcal{L} \sqcup \mathcal{C} \sqcup \mathcal{V}$, where
$\mathcal{L}$, $\mathcal{C}$ and $\mathcal{V}$ are learning, control and validation samples. Define

$$\mathbf{w}_{\{\bullet\}} = \arg \min_{i \in \{\bullet\} \subset \mathcal{I}} S(\mathbf{w}|\mathbf{X}_{\{\bullet\}}, \mathbf{y}_{\{\bullet\}}, f), \{\bullet\} = \mathcal{L}, \mathcal{C}, \mathcal{V}.$$

1) Regularity criterion: $S_{\Delta^2_{\mathcal{C}}} = \|\mathbf{y}_{\mathcal{C}} - \mathbf{X}_{\mathcal{C}}\mathbf{w}_{\mathcal{L}}\|^2$,
2) Criterion of Forecast skill:

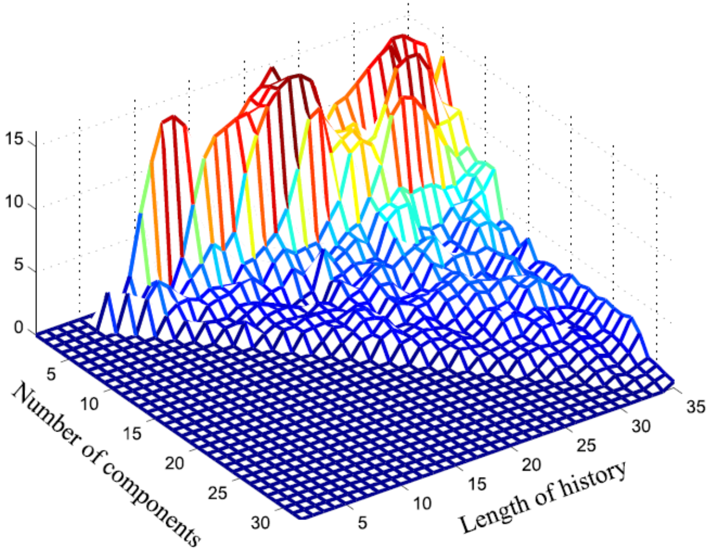
$$S_{\Delta^2_{\mathcal{V}}} = \frac{\|\mathbf{y}_{\mathcal{I}} - \mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{V}}\|^2}{\|\mathbf{y}_{\mathcal{I}} - \text{mean}(\mathbf{y}_{\mathcal{I}})\|^2},$$

3) Minimum-bias or consistency criterion:
$S_{\eta^2_{\mathbf{bs}}} = \|\mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{C}} - \mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{L}}\|^2$ or $S_{\eta^2_{\mathbf{a}}} = \|\mathbf{w}_{\mathcal{C}} - \mathbf{w}_{\mathcal{L}}\|^2$,
4) Robustness to noise

$$S_{V^2} = (\mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{C}} - \mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{I}})^{\mathsf{T}}(\mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{I}} - \mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{L}}) = (\mathbf{w}_{\mathcal{C}} - \mathbf{w}_{\mathcal{I}})^{\mathsf{T}}\mathbf{X}_{\mathcal{I}}^{\mathsf{T}}\mathbf{X}_{\mathcal{I}}(\mathbf{w}_{\mathcal{I}} - \mathbf{w}_{\mathcal{L}}),$$

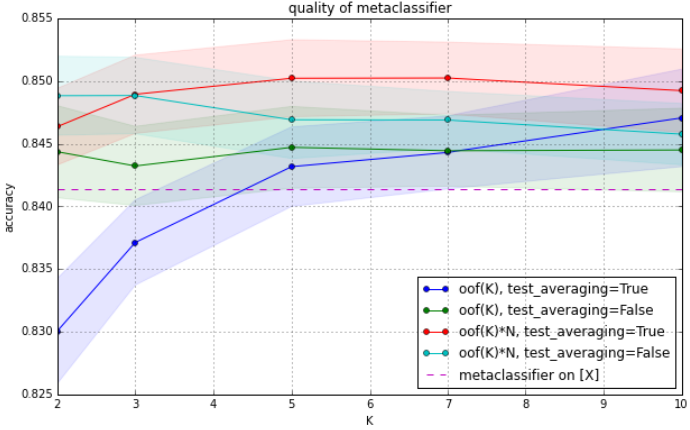
5) Combined criterion: $S_{\kappa^2} = \sum\limits_{k=1}^{K} v_k S_k$ with $\sum\limits_{k=1}^{K} v_k = 1$.

jmlda01



15
10
5
0
5
10
15
20
25
30
Number of components
5
10
15
20
25
30
35
Length of history

guschin


quality of metaclassifier
0.855
0.850
0.845
0.840
0.835
0.830
0.825
accuracy
2
3
4
5
6
7
8
9
10
K
oof(K), test_averaging=True
oof(K), test_averaging=False
oof(K)*N, test_averaging=True
oof(K)*N, test_averaging=False
metaclassifier on [X]