ВШ ИСКУСТВЕННОГО ИНТЕЛЛЕКТА

курс: Архитектура суперкомпьютерных систем

ЛЕКЦИЯ 2: ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ДИСПЕТЧЕРСКОГО УПРАВЛЕНИЯ ГИБРИДНЫМИ ВЫСОКОПРОИЗВОДИТЕЛЬНЫМИ КЛАСТЕРАМИ ВЫЧИСЛИТЕЛЬНЫМИ КЛАСТЕРАМИ.

12.09.2023

Содержание

- **Введение**. Что мы хотим от «Машинного обучения» высокопроизводительных вычислительных кластеров
- **Возможности** повышения «реальной производительности» гибридных СК кластеров:
  – интеллектуальное диспетчерское управление
- **Оценка** параметров, используемых диспетчером slurm, на основе машинного обучения модели «выживаемости» прикладных задач
- **Результаты** моделирования
- **Выводы**

Введение. Что мы хотим от «машинного обучения» СК

- **Имеем:** в настоящее время «машинное обучение» – стало синонимом «больших данных». Без гигантских объемов заранее размеченных данных алгоритмы машинного обучения не способны извлечь абстрактные знания, которые затем можно перенести на новые ситуации

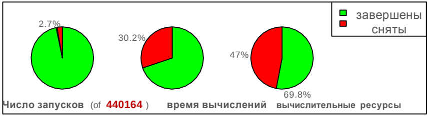
- **Вопрос:** как получить максимум информации из минимума данных, то есть выявить суть явления или процесса из наблюдений ?
- **Ответ**: нужно **встроить в СК искусственный механизм внимания**, чтобы … эффективно
  - фильтровать поступающую информацию,
  - выявлять релевантные связи,
  - строить гипотезы и обобщать полученные данные,

  чтобы развивать возможности оценивать вероятности соответствия выбранных моделей и гипотез результатам наблюдений, извлекая из «сырых» данных объяснительные абстракции (функция выживания), которые можно переносить на другие ситуации ……

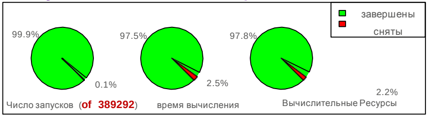
- **Как это сделать.** Для большинства алгоритмов машинного обучения настраиваемые параметры – это веса нейронов или ветвей «случайных деревьев»…, поэтому задача «обучения» состоит в поиске такого сочетания всех параметров модели процесса , которые вместе определят модель «наилучшего из возможных решений».

Возможности «умных» вычислений

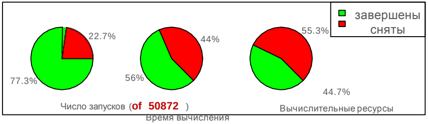
Общая **характеристика эффективности для всех видов запусков заданий**



Класс эквивалентности 1: запуск заданий на СК в «ИИ режиме»



Класс эквивалентности 2: запуск неизвестных задания «в ручном" режиме



Вывод:

Если параметры заданий пользователей, хорошо (точно) известны, то настройки процессов вычислений можно автоматизировать и более эффективно использовать имеющие компьютерные ресурсы

Как повысить реальную производительность суперкомпьютера ?

Есть два основных направления:

\

1) Разрабатывать новые компьютерные архитектуры, ППО и электронную компонентную базу

2) Построить модель задачи, которая на основе :

- используется для проведения прямого вычисления правильного ответа на основе заданных начальных данных и «формулы» описания процесса, реализуемого на имеющейся аппаратно-программной базе СК
- «опыта» научиться передсказывает правильный ответ для решаемой задачи с вероятностью 0.9Х
- использования методов «машинного обучения» объяснит как надо распределить имеющие ресурсы кластера между всеми прикладными задачами, чтобы с вероятностью 0.9Х они завершилась успешно, объяснив при этом, почему выбран именно такой вариант действий

Цель обучения – повышение реальной производительности

Реальную производительность работы гибридного СКЦ, работающего в режиме центра коллективного пользования (ЦКП, можно повысить на основе такого **планирования процесса** распределения ресурсов гибридного кластера, которое обеспечивает успешное завершение задачи пользователя за интервал времени, выделенный для нее диспетчером slurm

Реальная производительность работы СКЦ оценивается через:

- Вероятность успешного (или не успешного) выполняется прикладной задачи за выделенное slurm время выполнения,
- Количество успешно выполненных задач пользователей за заданный интервал времени

Для этого требуются оценки параметров прикладных заданий и самого СК, которые используются для системой планирования slurm и влияют на реальную производительность работы СКЦ, однако могут быть неизвестны пользователю на момент запуска задачи.

Для **системы машинного обучения СКЦ** используется цензурированная выборка данных, включая

- $x_i \in \mathcal{X}$ – входной вектор параметров задания, задаваемых непосредственно самим пользователем
- $t_i \in [0, +\infty)$ – время до события (завершения задачи)
- $\delta_i \in \{0, 1\}$ – <u>индикатор цензурирования</u> (1, если задача завершена до истечения времени, 0 иначе)
- $y_i \in \mathcal{Y}$ – вектор параметров задачи, не известный на момент запуска (например, факторы изменения состояния системы slurm), в который могут входить <u>индикаторы ошибки выполнения задачи</u> (0 – успешное завершение, 1 – завершение с ошибкой)

Формулировка задачи обучения диспетчера СК:

Построить модель оценки параметров прикладного задания (оценку ожидаемого времени решения) **с** использованием **механизмов внимания** в условиях цензурирования данных, а именно:

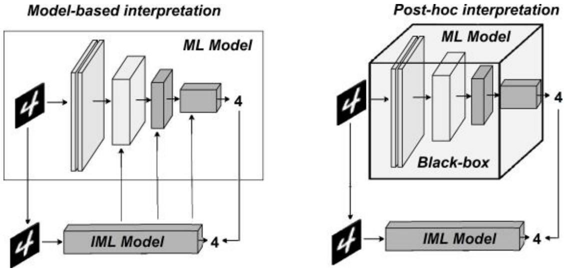
- Ожидаемое время решения $\mathbb{E}[T|X = x]$
- Функции выживаемости прикладной задачи в среде СК, а именно S(t, X)
- вектора параметров $\hat{y}$ выполнения задачи, компонентами которого являются вероятность успешного выполнения задания P(X=x)

С целью использования этих параметров в системе планирования ресурсов СКЦ, построенной на базе **диспетчера slurm**

Post-hoc и model-based интерпретация результатов обучения

Post-hoc интерпретация - объяснение после обучения основной модели, пример - метод LIME (Local Interpretable Model-Agnostic Explanations)

Model-based интерпретация результата, которая позволяет "вмешиваться" в процесс обучения и повысить «вероятность» получения правильного ответа



интерпретация: локальная и глобальная

1. Локальная - методы интерпретации относятся к отдельному примеру и результатам его классификации, а именно:

- - Почему модель принимает конкретное решение для данной задачи?
  - Почему модель принимает конкретные решения для группы задач одного класса?

1. Глобальная - методы интерпретации сфокусированы на значимых признаках всего множества анализируемых задач

Модель задачи интерпретация полученных решений

Любая интерпретация – это «решение обратной задач», которая имеет множество решений, поэтому нуждается в регуляризации.

Формально модель задачи это функция $f : R^m \rightarrow R^D$, например, в задаче классификации $f(\mathbf{x})$ - это вероятность того, что $\mathbf{x}$ принадлежит определенному классу

Интерпретация - это также модель $g(x) \in G$, где G - класс эффективно интерпретируемых моделей, например:

линейная регрессия, деревья решений,…,случайные леса

Задача оптимизации модели интерпретации:

$$\min_{g \in G} \{L(f, g, \theta) + \Omega(g)\}$$

$L(f, g, \theta)$ - мера того, как неточна g в аппроксимации f

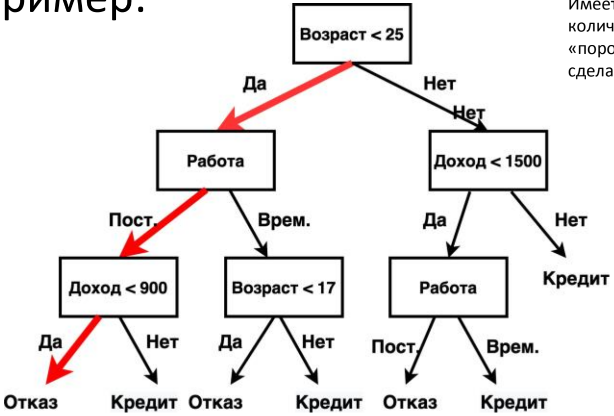
θ - вектор параметров; $\Omega(g)$ - регуляризатор

Классы интерпретируемых моделей

- Линейная регрессия
- Логистическая регрессия
- Деревья решений
- GLM (Лассо, гребневая регрессия, эластичная сеть)
- К ближайших соседей

Выбор объяснения на основе деревья решений, почему ?

пример:

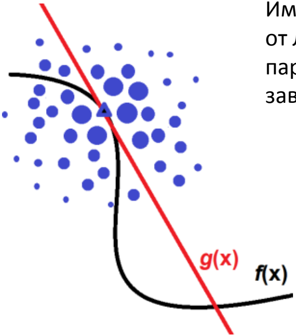


Имеется количественный «порог» обоснования сделанного выбора

Выбор объяснения на основе регрессии, почему ?

Линейная регрессия: $g(\mathbf{x}) = a_1x_1 + a_2x_2 + \ldots + a_mx_m$

GAM: $g(\mathbf{x}) = g_1(x_1) + g_2(x_2) + \ldots + g_m(x_m)$



Имеется линейная зависимость от локально измеряемых параметров или линейная зависимоть от функций от этих

Модели "выживаемости" прикладной задачи в СКЦ

Дано: Рассматривается «событие» - вероятность «успешного решения» задач пользователя до истечения заданного диспетчером времени Ti, которая связано с вектором $\mathbf{x}$. **Имеется** обучающая выборка $(\mathbf{x}_i, \delta_i, T_i)$, где вектор $\mathbf{x}_i = (x_{i1}, \dots, x_{id})$, $i = 1, \dots, n$ - цензурированные данные о функционировании СК во время выполнения задания, включая параметры прикладной задачи, которые задаются самим пользователем

Заданы функции

- риска $h(t|\mathbf{x})$, что задача завершиться в момент времени t
- кумулятивная функция риска $H(t|\mathbf{x})$,
- функция выживаемости $S(t|\mathbf{x}) = \Pr\{T > t|\mathbf{x}\}$

(пример: модель пропорционального риска Кокса: $H(t|\mathbf{x}, \mathbf{b}) = H_0(t) \exp(\mathbf{x}\mathbf{b}^T)$

Используемые модели выживаемости

Рассматривается модель выживаемости на основе случайного леса и обучающая выборка $(\mathbf{x}_i, \delta_i, T_i)$, $i = 1, ..., n$.

Выход модели либо функция

- риска
- выживаемости (имеет не одно значение)

Требуется определить какие компоненты вектора $\mathbf{x}$ определяют функцию риска $H(t|\mathbf{x})$?

Например, функция риска на основе модели Кокса имеет вид:

$$H_{\mathrm{Cox}}(t|\mathbf{x}, \mathbf{b}) = H_0(t) \exp\left(\mathbf{x}\mathbf{b}^{\mathrm{T}}\right)$$

Модель SurvLIME детально

Роль выхода модели прогноза $y = f(\mathbf{x})$ (класса) играет кумулятивная функция риска $H(t|\mathbf{x})$,

Функцию риска $H(t|\mathbf{x})$ будем аппроксимировать моделью Кокса

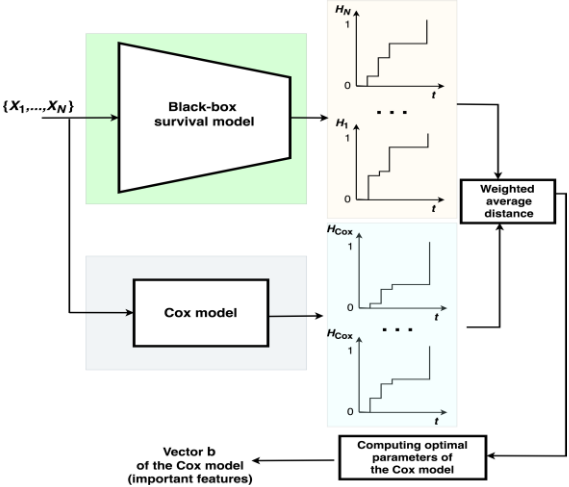
$H_{\text{Cox}}(t|\mathbf{x}, \mathbf{b}) = H_0(t)\exp\left(\mathbf{x}\mathbf{b}^{\mathrm{T}}\right)$, чтобы получить зависимость

$H_{\text{Cox}}(t|\mathbf{x}, \mathbf{b})$ от $\mathbf{x}$ с коэффициентами $\mathbf{b}$

Для этого сгенерируем «облако» точек $\mathbf{x}_k$ вблизи $\mathbf{x}$ с весами $w_k = K(\mathbf{x}_k, \mathbf{x})$

Оптимальные значения $\mathbf{b}$ определяются минимизируя среднее "весовое" расстояние между $H(t|\mathbf{x})$ и $H_{\text{Cox}}(t|\mathbf{x}, \mathbf{b})$

Иллюстрация SurvLIME



Задача оптимизации для вычисления коэффициентов

$$\min_{\mathbf{b}} \sum_{k=1}^{N} w_k \sum_{j=0}^{m} v_{kj}^2 \left(\ln H_j(\mathbf{x}_k) - \ln H_{0j} - \mathbf{b}^{\mathrm{T}} \mathbf{x}_k \right)^2 (t_{j+1} - t_j),$$

где

$$w_k = 1 - \sqrt{\frac{\| \mathbf{x} - \mathbf{x}_k \|_2}{r}}, \quad v_{kj} = H_j(\mathbf{x}_k) / \ln(H_j(\mathbf{x}_k)).$$

$\mathbf{x}_k$ - сгенерированные точки, $\mathbf{b}$ - вектор искомых коэффициентов

$H_0(\mathbf{x}_k)$, $H(\mathbf{x}_k)$ - базовая и обычная кумулятивные функции риска

N - количество сгенерированных точек

Числовые эксперименты с синтетическими данными

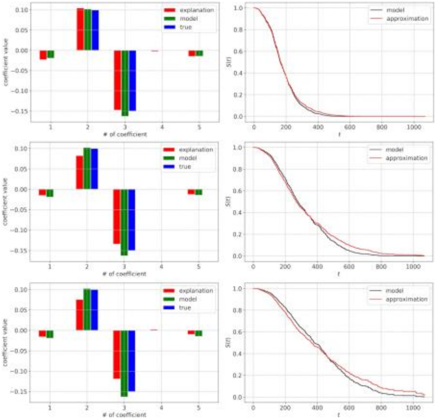


Рис.: Наилучшая, средняя и наихудшая аппроксимации для модели Кокса как черного ящика

Числовые эксперименты с реальными данными

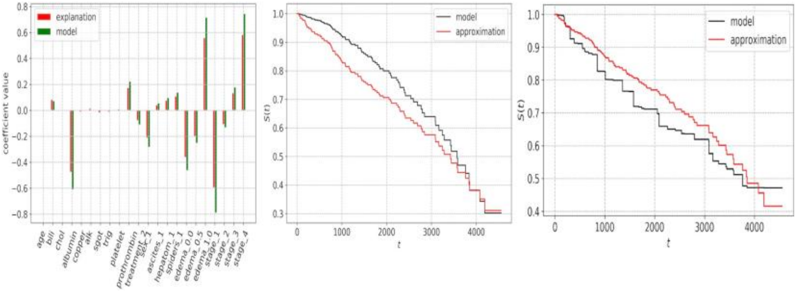


Рис.: Аппроксимации для модели Кокса как черного ящика (первый и второй графики) и случайного леса выживаемости (третий график), обученных на датасете Primary Biliary Cirrhosis

Числовые эксперименты с реальными данными

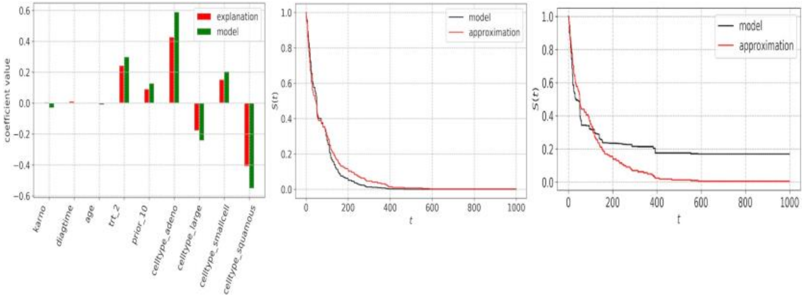


Рис.: Аппроксимации для модели Кокса как черного ящика (первый и второй графики) и случайного леса выживаемости (третий график), обученных на датасете Veterans' Administration Lung Cancer Study

Модели выживаемости, учитывающие совместное влияние признаков

Имеется основная обученная модель выживаемости (случайный лес выживаемости, нейронная сеть) и датасет $(\mathbf{x}_i, \delta_i, T_i)$, $i = 1, ..., n$.

Вопрос: какие пары признаков некоторого вектора $\mathbf{x}$ совместно определяют его функцию риска $H(t|\mathbf{x})$?

Идея: использовать метод Лассо и произведение признаков в расширенной модели Кокса:

$$H_{\text{Cox}}(t\,|\mathbf{x}, \mathbf{z}, \mathbf{b}, \mathbf{c}) = H_0(t) \exp\left(\mathbf{x}\mathbf{b}^{\mathrm{T}} + \mathbf{z}\mathbf{c}^{\mathrm{T}}\right)$$

где $\mathbf{z}$ - вектор произведений пар признаков $\mathbf{z} = (x_1 x_2, x_1 x_3, ..., x_{d-1} x_d)$;
$\mathbf{c}$ - искомый вектор коэффициентов длины $d(d-1)/2$:
$\mathbf{c} = (c_{1,1}, c_{1,2}, ..., c_{d-1,d})$

Более детально

1. Анализ аналогичен SurvLIME, где функция риска $H(t|\mathbf{x})$ - выход основной модели "черного ящика".
2. Аппроксимируем $H(t|\mathbf{x})$ функцией риска расщиренной модели Кокса $H_{\mathrm{Cox}}(t|\mathbf{x},\mathbf{z},\mathbf{b},\mathbf{c}) = H_0(t)\exp\left(\mathbf{x}\mathbf{b}^{\mathrm{T}} + \mathbf{z}\mathbf{c}^{\mathrm{T}}\right)$
3. Для построения модели Кокса генерируем много точек $\mathbf{x}_k$ вблизи $\mathbf{x}$ с весами $w_k = K(\mathbf{x}_k,\mathbf{x})$
4. Оптимальные значения $\mathbf{b}$ определяются минимизируя среднее весовое расстояние между $H(t|\mathbf{x})$ и $H_{\mathrm{Cox}}(t|\mathbf{x},\mathbf{z},\mathbf{b},\mathbf{c})$ с учетом **регуляризации**

Задача оптимизации для вычисления коэффициентов

$$\min_{\mathbf{b},\mathbf{c}} \sum_{k=1}^{N} w_k \sum_{j=0}^{m} v_{kj}^2 \left(\ln H_j(\mathbf{x}_k) - \ln H_{0j} - \mathbf{b}^{\mathrm{T}} \mathbf{x}_k - \mathbf{c}^{\mathrm{T}} \mathbf{z}_k \right)^2 (t_{j+1} - t_j)$$

$$+ \lambda \sum_{i=1}^{d} \sum_{j=1, j/=i}^{d} |c_{i,j}| + \mu \sum_{i=1}^{d} b_i^2.$$

Слагаемое $\lambda \sum_{i=1}^{d} \sum_{j=1,j/=i}^{d} |c_{i,j}|$ - регуляризационное слагаемое в соответствии с методом Лассо, используется для ограничения множества ненулевых коэффициентов в векторе $\mathbf{c}$.

Регуляризация вектора $\mathbf{b}$ ($\mu \sum_{i=1}^{d} b_i^2$) осуществляется с использованием квадратичной нормы.

Границы Колмогорова-Смирнова

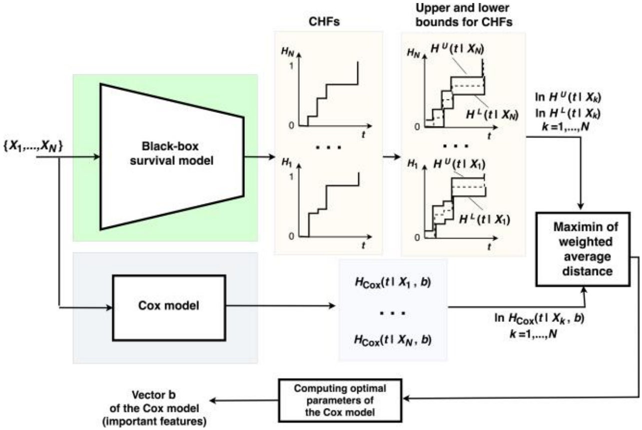
$$F_n^L(\mathbf{x}) \leq F(\mathbf{x}) \leq F_n^U(\mathbf{x}),$$

$$F_n^L(\mathbf{x}) = \max(F_n(\mathbf{x}) - d_{n,1-\gamma}, 0),$$
$$F_n^U(\mathbf{x}) = \min(F_n(\mathbf{x}) + d_{n,1-\gamma}, 1).$$

Для $n > 10$: $d_{n,1-\gamma} \approx k_{1-\gamma}/\sqrt{n}$

Для $n \leq 10$: $d_{n,1-\gamma} \approx k_{1-\gamma}\left(\sqrt{n} + 0.12 + 0.11/\sqrt{n}\right)^{-1}$

пример SurvLIME-KS



задача максиминной оптимизации

$$\max_{\grave{}_k} \min_{\mathbf{b}, z_k} L(\mathbf{b}) = \max_{\grave{}_k} \min_{\mathbf{b}, z_k} \sum_{k=1}^{N} w_k z_k ,$$

при ограничениях $\theta_k \in F_k, k = 1, ..., N,$

$$z_k \geq \max_{j=0,...,m} \theta_{kj} - \mathbf{x}_k \mathbf{b}^{\mathrm{T}}, \ k = 1, ..., N,$$

$$z_k \geq \mathbf{x}_k \mathbf{b}^{\mathrm{T}} - \min_{j=0,...,m} \theta_{kj}, \ k = 1, ..., N$$

F_k - множество θ_k, определяемое из

$$\theta^L_{k,j} \leq \theta_{k,j} \leq \theta^U_{k,j}, \ k = 1, ..., N, \ j = 0, ..., m,$$

$$\theta_{k,j} \leq \theta_{k,j+1} + \ln(H_{0j+1}/H_{0j}), \ j = 0, ..., m-1$$

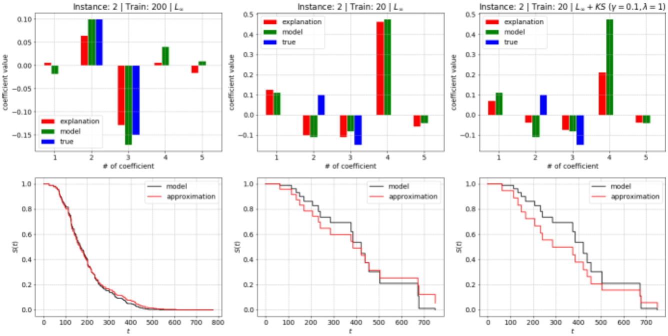
SurvLIME-KS и задача квадратичного программирования

$$\min_{\mathbf{b}} \sum_{k=1}^{N} w_k z_k + \lambda \|\mathbf{b}\|^2,$$

$$z_k + \mathbf{x}_k \mathbf{b}^{\mathrm{T}} \geq \max_{j=0,\dots,m} \theta^U_{k,j}, \; k = 1, \dots, N,$$

$$z_k - \mathbf{x}_k \mathbf{b}^{\mathrm{T}} \geq - \min_{j=0,\dots,m} \theta^L_{k,j}, \; k = 1, \dots, N.$$

Числовые эксперименты с синтетическими данными



Числовые эксперименты с реальными данными

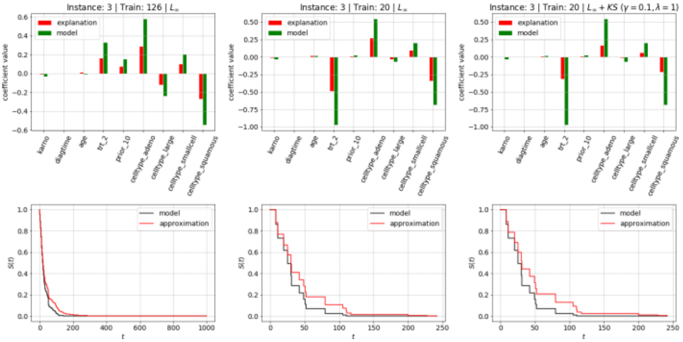
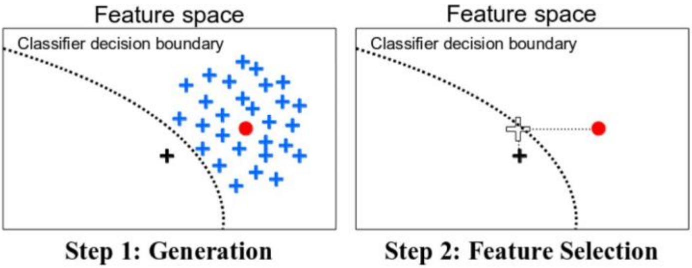


Рис.: An illustration of the approximation results under three conditions of experiments for the lack- ox Cox model trained on the Veteran dataset

Counterfactual объяснения (гипотетические, противопоставления)

- Counterfactual - наименьшие изменения значений признаков, которые изменяют класс примера
- "Ваш запрос на кредит отклонен, так как ваш доход $30,000 и ваш баланс $200. Если бы ваш доход был $35,000 и ваш текущий баланс был $400, то ваш запрос был бы одобрен"



Конфликтные примеры (counterfactual)

f - функция 11черного ящика11. Counterfactual пример $\mathbf{z} = (z_1, ..., z_d) \in \mathrm{R}^d$ для объясняемого примера $\mathbf{x} = (x_1, ..., x_d) \in \mathrm{R}^d$ определяется из решения задачи оптимизации:

$$\min_{\mathbf{z} \in \mathrm{R}^m} L(f(\mathbf{z}), f(\mathbf{x})) + C\theta(\mathbf{z}, \mathbf{x}),$$

где $L(\cdot, \cdot)$ - функция потерь, устанавливающая отношение между предсказаниями модели f;
$\theta(\cdot, \cdot)$ - штрафное слагаемое для отклонения $\mathbf{z}$ от $\mathbf{x}$, например евклидово расстояние;
$C > 0$ - параметр

Конфликтные примеры (counterfactual)

Постановка задачи для модели выживаемости

$$m_{\mathbf{z}} = \mathrm{E}(\mathbf{z}) = \int_0^{\infty} S(t \mid \mathbf{z})\mathrm{d}t, \; m_{\mathbf{x}} = \mathrm{E}(\mathbf{x}) = \int_0^{\infty} S(t \mid \mathbf{x})\mathrm{d}t.$$

Задача оптимизации:

$$\min_{\mathbf{z} \in \mathbb{R}^d} L(m_{\mathbf{z}}, m_{\mathbf{x}}) + C\theta(\mathbf{z}, \mathbf{x}).$$

Ограничение на классы:

$$m_{\mathbf{z}} - m_{\mathbf{x}} \geq r.$$

Конфликтные примеры (counterfactual)

Итоговая задача оптимизации

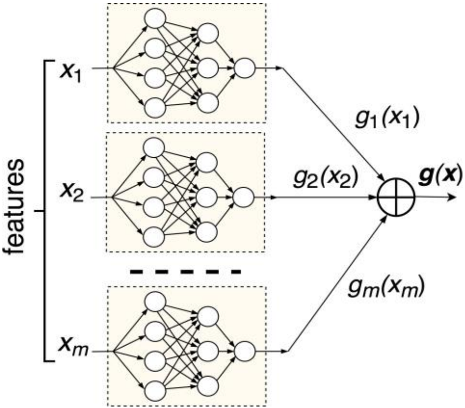
$$\min_{\mathbf{z}\in\mathbb{R}^m} L(\mathbf{z}) = \min_{\mathbf{z}\in\mathbb{R}^m} \{\max\{0, C\psi(\mathbf{z})\} + \|\mathbf{z}-\mathbf{x}\|_2\},$$

где

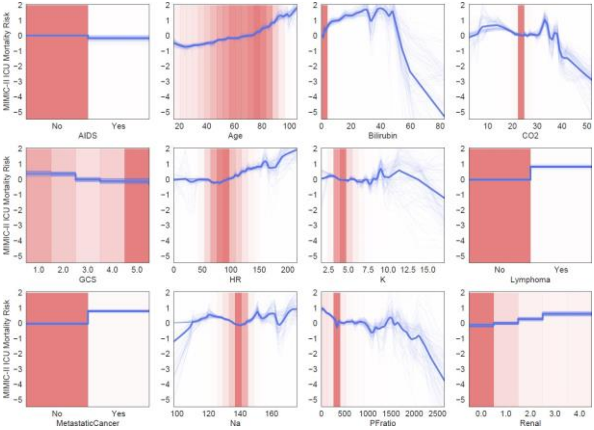
$$\psi(\mathbf{z}) = r - (m(\mathbf{z}) - m(\mathbf{x})) \leq 0.$$

NAME (neural additive model)

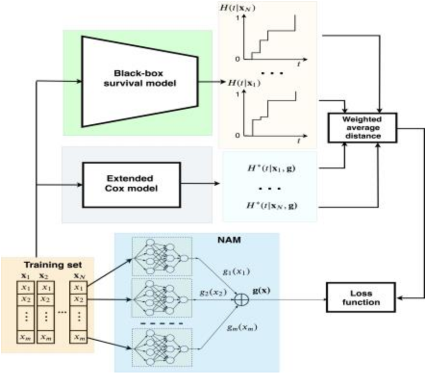
GAM: $g(\mathbf{x}) = g_1(x_1) + g_2(x_2) + \ldots + g_m(x_m)$



NAM: результаты (графики частичной зависимости)

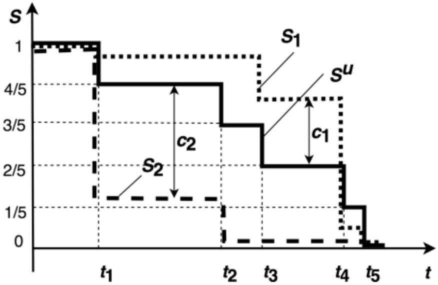


SurvNAM



$$L(\mathbf{W},D)=\sum_{i=1}^{N} K(\mathbf{x},\mathbf{x}_i)\sum_{j=0}^{s} \tau_j \left(\ln H_j(\mathbf{x}_i) - \ln H_{0j} - \sum_{k=1}^{m} g_k(x_k^{(i)})\right)^2$$

Неопределенность предсказаний



Неопределенность предсказаний - объяснение