



Физический факультет
Московского
государственного университета
имени М.В.Ломоносова

**ОПТИМИЗАЦИЯ ПРЕДОБРАБОТКИ ВХОДНЫХ
ПРИЗНАКОВ ПРИ ПРОГНОЗИРОВАНИИ
ГЕОМАГНИТНОГО ИНДЕКСА DST МЕТОДАМИ
МАШИННОГО ОБУЧЕНИЯ**
**OPTIMIZATION OF PREPROCESSING OF INPUT FEATURES
IN PREDICTING THE GEOMAGNETIC DST INDEX USING
MACHINE LEARNING METHODS**

Курсовая работа студента 2 курса Сухадольского А.Г.

Научные руководители:

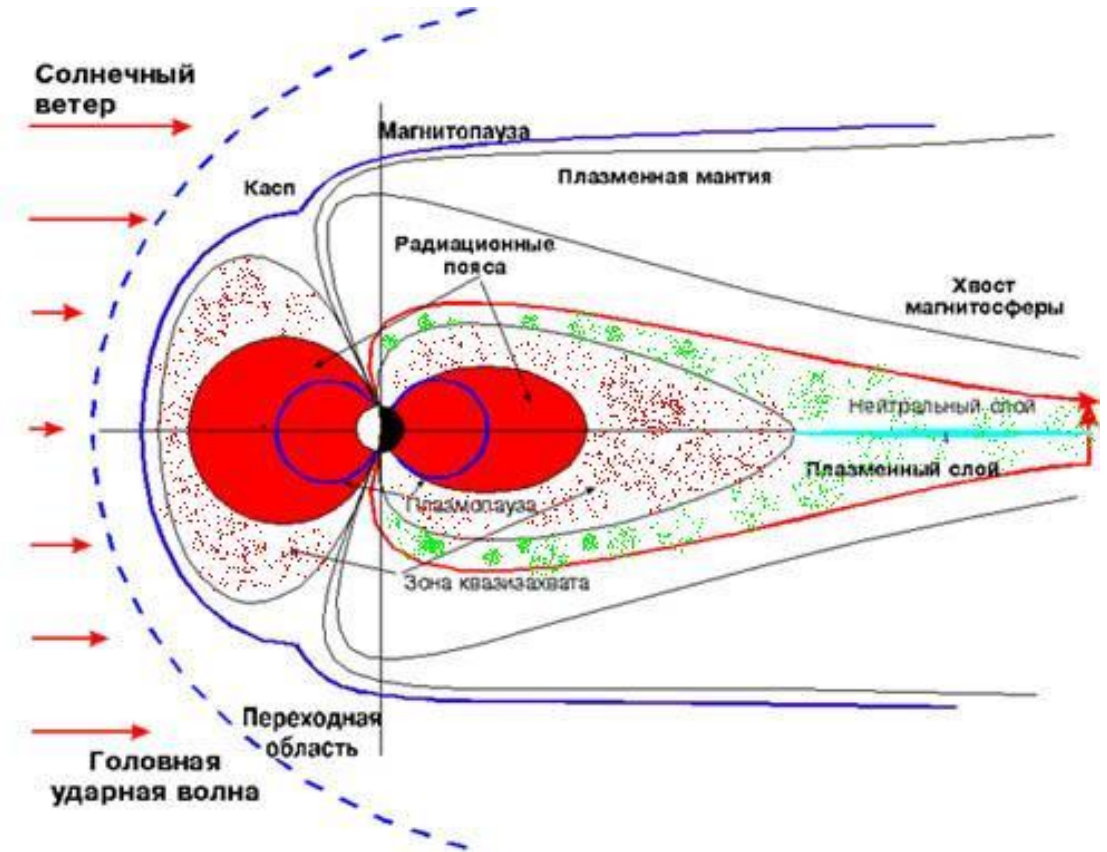
Зав. лаб., к.ф.-м.н. Доленко С. А.

Инженер Владимиров Р.Д.

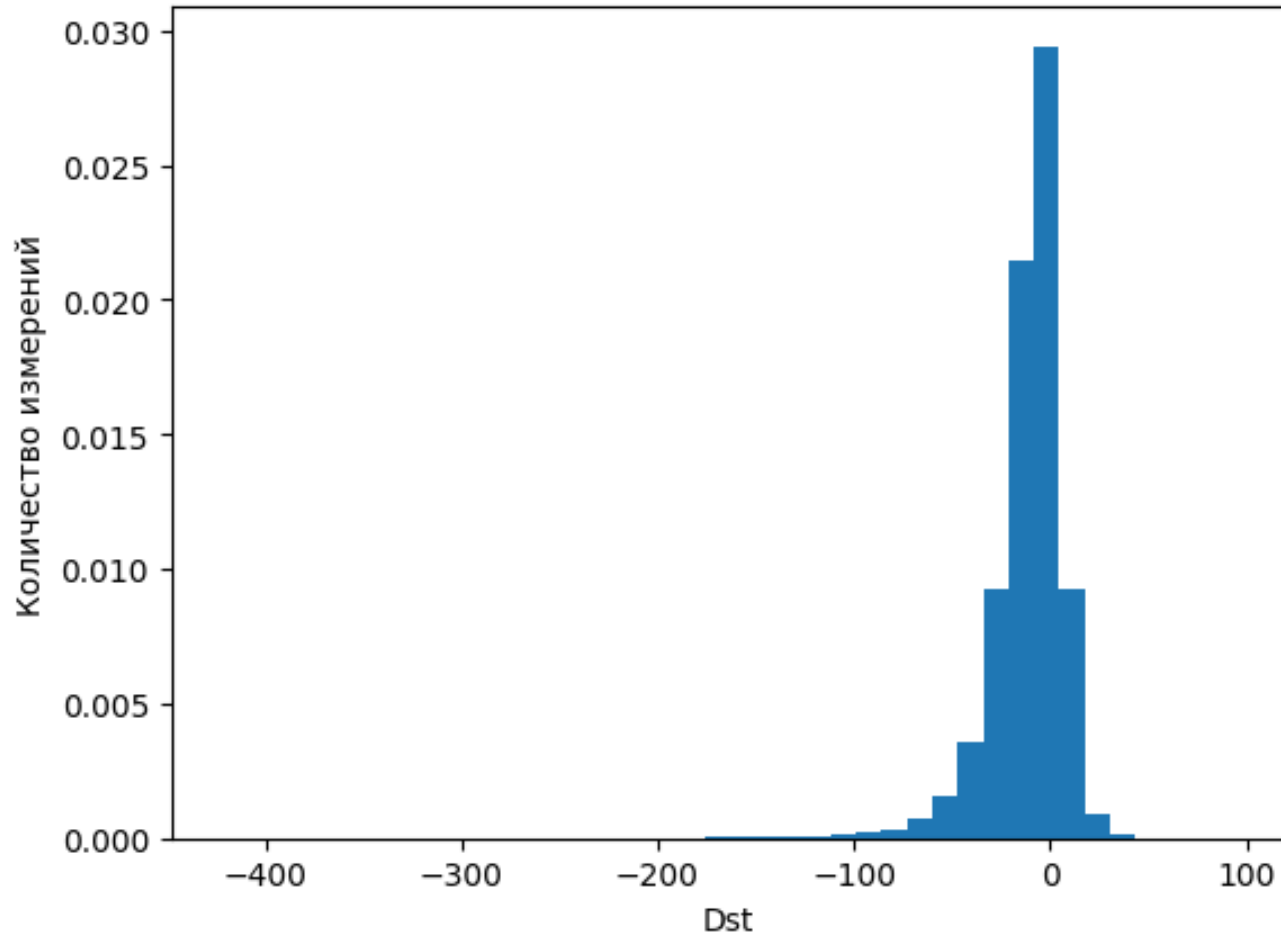


Введение.

Магнитным полем Земли или геомагнитным полем называют поле, генерируемое внутриземными источниками. Оно испытывает возмущения, называемые геомагнитными вариациями, при взаимодействии его внешнего слоя, магнитосферы, с солнечным ветром - потоком плазмы, электронов и ионов, преимущественно протонов, из солнечной короны.



индекс Dst



В данной работе взят за основу индекс шторм-тайм вариации (storm-time variation) Dst который представляет собой максимальное отклонение от спокойного уровня на контрольной цепочке магнитных станций. Dst-индекс вычисляется один раз за час. Это один из наиболее часто используемых индексов при исследовании космической погоды и её влияния на магнитосферу Земли.



Данные

В качестве входных данных использовались значения следующих физических величин:

- ♦ **Скорость солнечного ветра в точке Лагранжа L1 (км/с)**
- ♦ **Плотность протонов в солнечном ветре в точке Лагранжа L1 (см⁻³)**
- ♦ **Векторные компоненты межпланетного магнитного поля и модуль вектора магнитной индукции (нТ)**

Было произведено так называемое погружение многомерного временного ряда, состоящего из перечисленных величин, на 24 часа, т.е. массив входных признаков был дополнен значениями каждой из этих величин за предыдущие 24 часа. Погружение производилось как с обычными данными, так и с предварительно логарифмированными



Цель и задачи работы.

Цель данной работы - проверка влияния на качество прогнозирования Dst-индекса предобработки его значений с помощью нелинейного преобразования.

Задачами, решаемыми для достижения данной цели, были:

- 1) Программная **реализация обучения и применения искусственной нейронной сети** типа многослойный персептрон.
- 2) **Исследование результатов прогнозирования индекса Dst с горизонтом от 1 до 24 часов** с помощью линейной регрессии и МСП.
- 3) **Исследование результатов прогнозирования индекса Dst с использованием тех же ММО с предварительным нелинейным преобразованием его значений путём логарифмирования**



Эксперимент

Были использованные:

- ◆ Линейные модели:

 - ◆ Линейная модель без регуляризаций

 - ◆ Регуляризация лассо (L1)

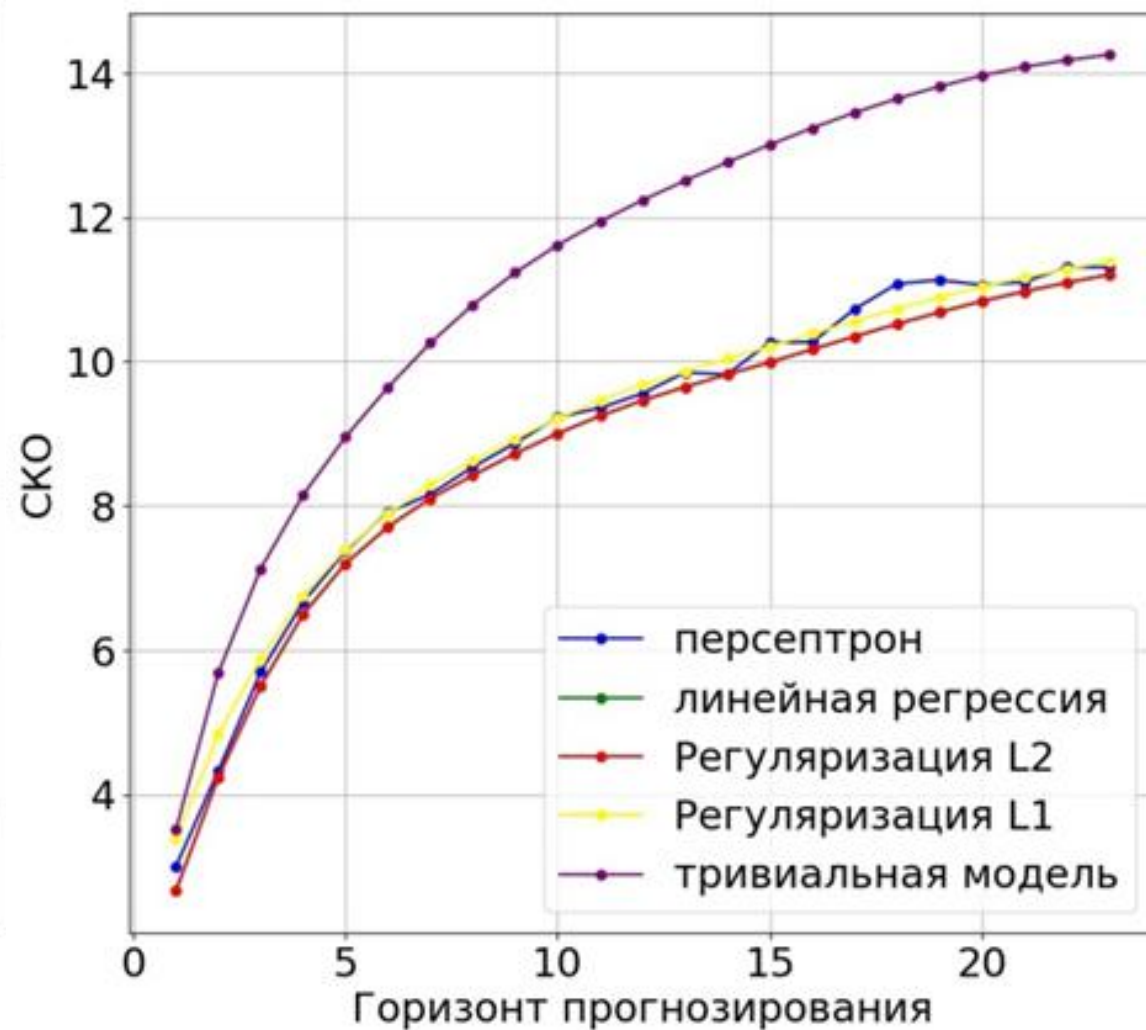
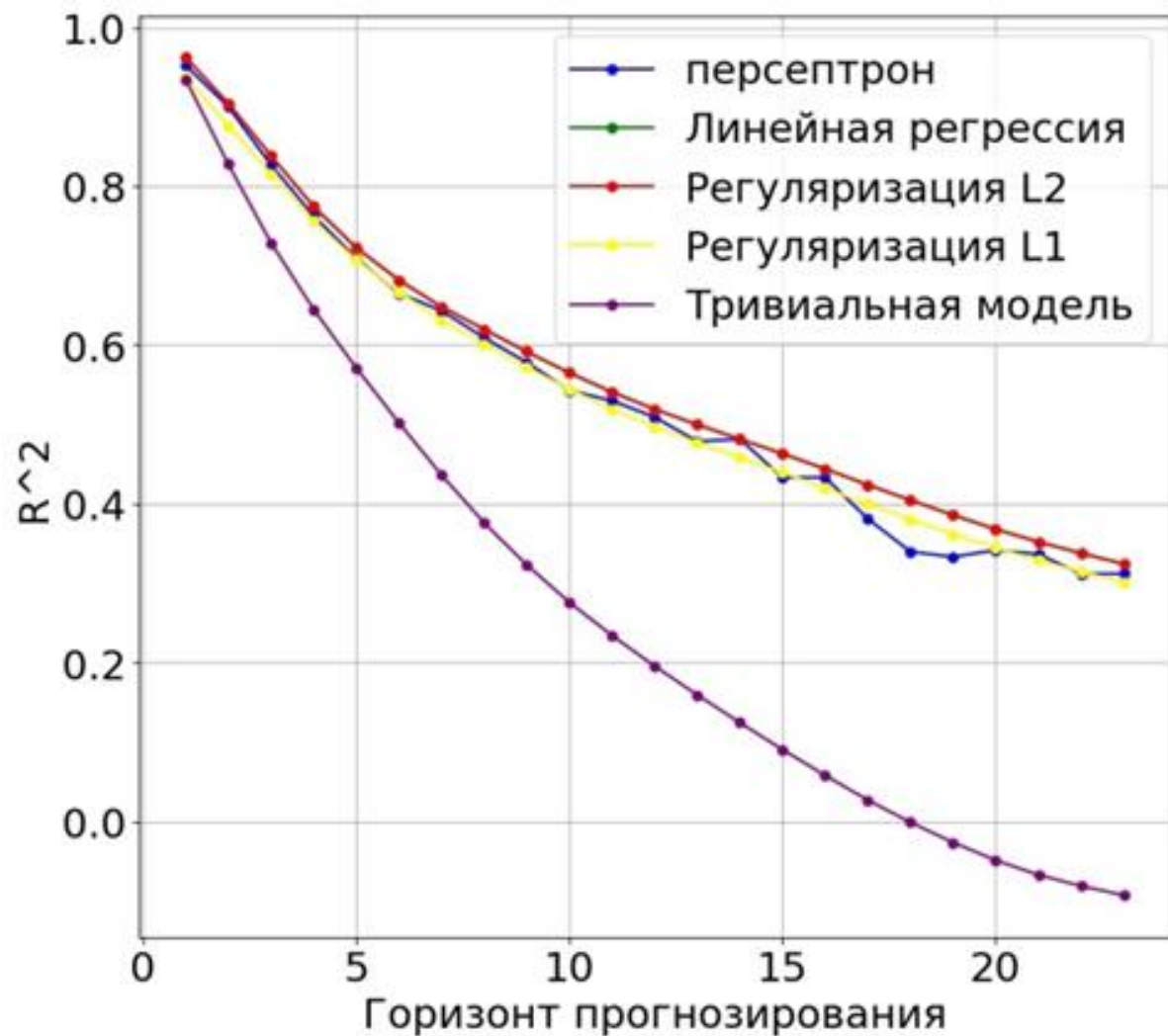
 - ◆ Гребневая регрессия (L2)

- ◆ многослойный персептрон с одним скрытым слоем из 64 нейронов. На вход персептрона поступает 148 параметров.

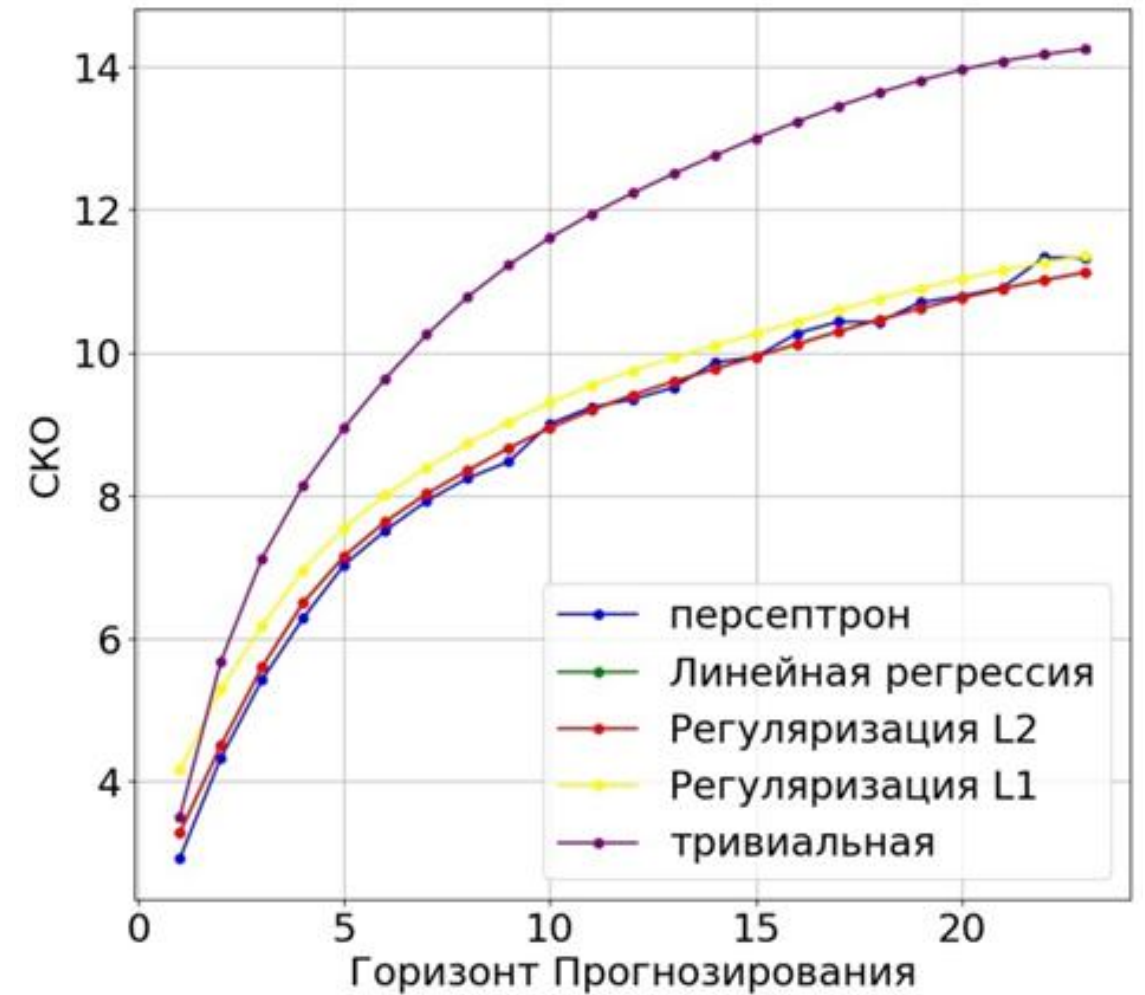
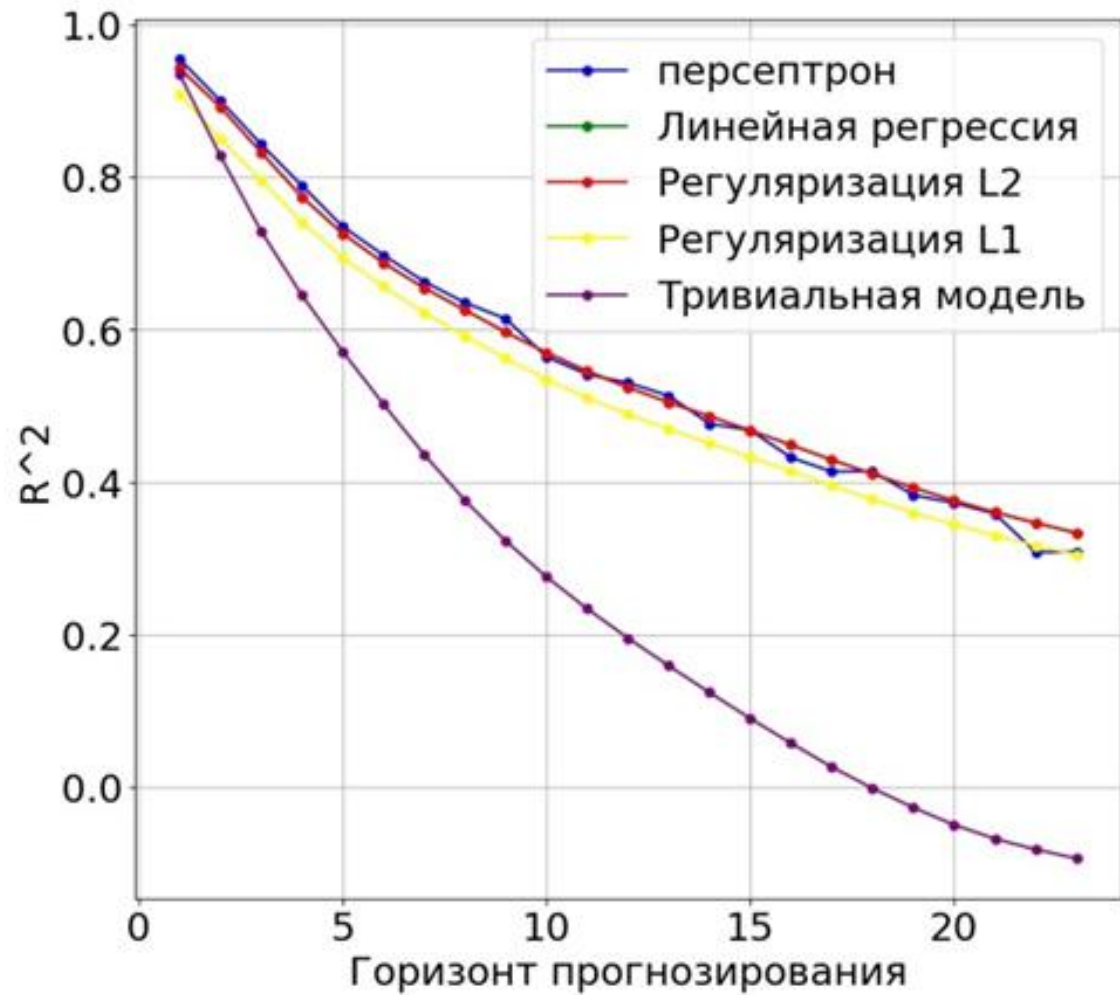


Зависимость качества прогнозирования от горизонта прогноза.

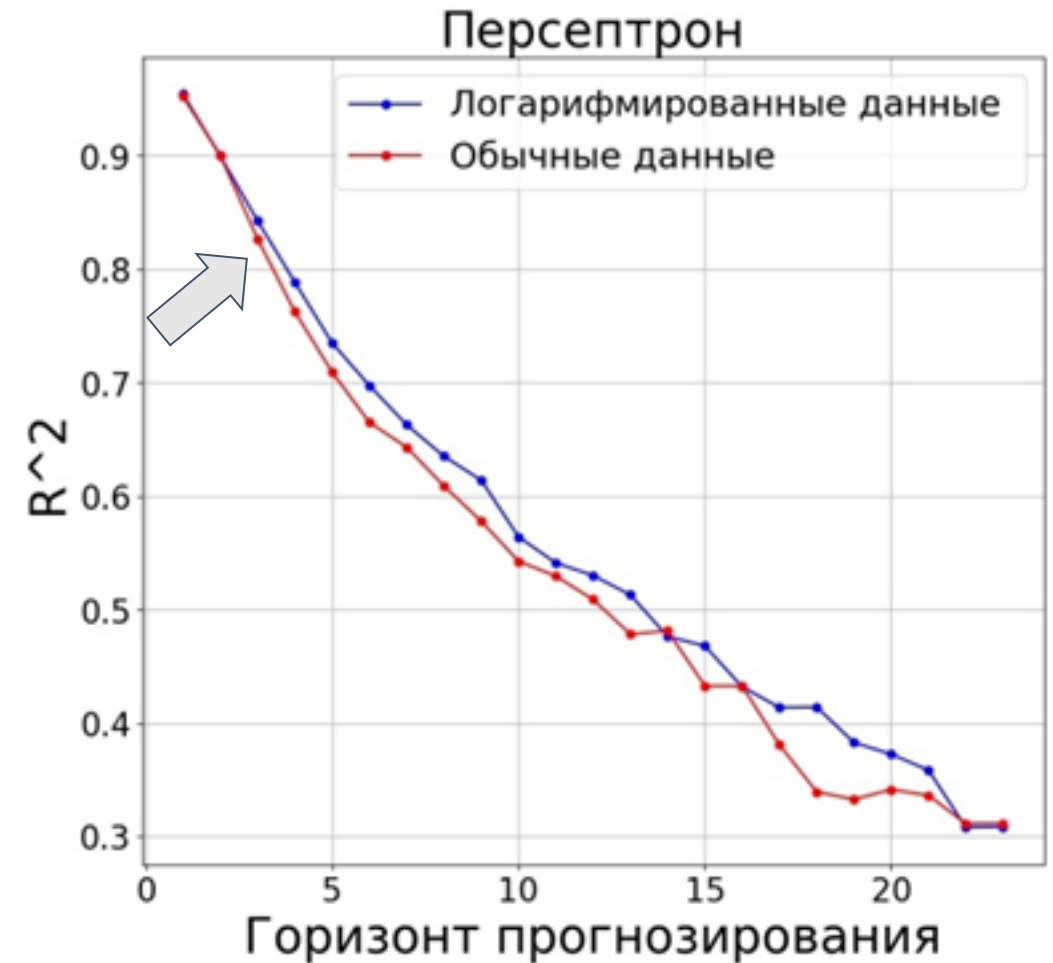
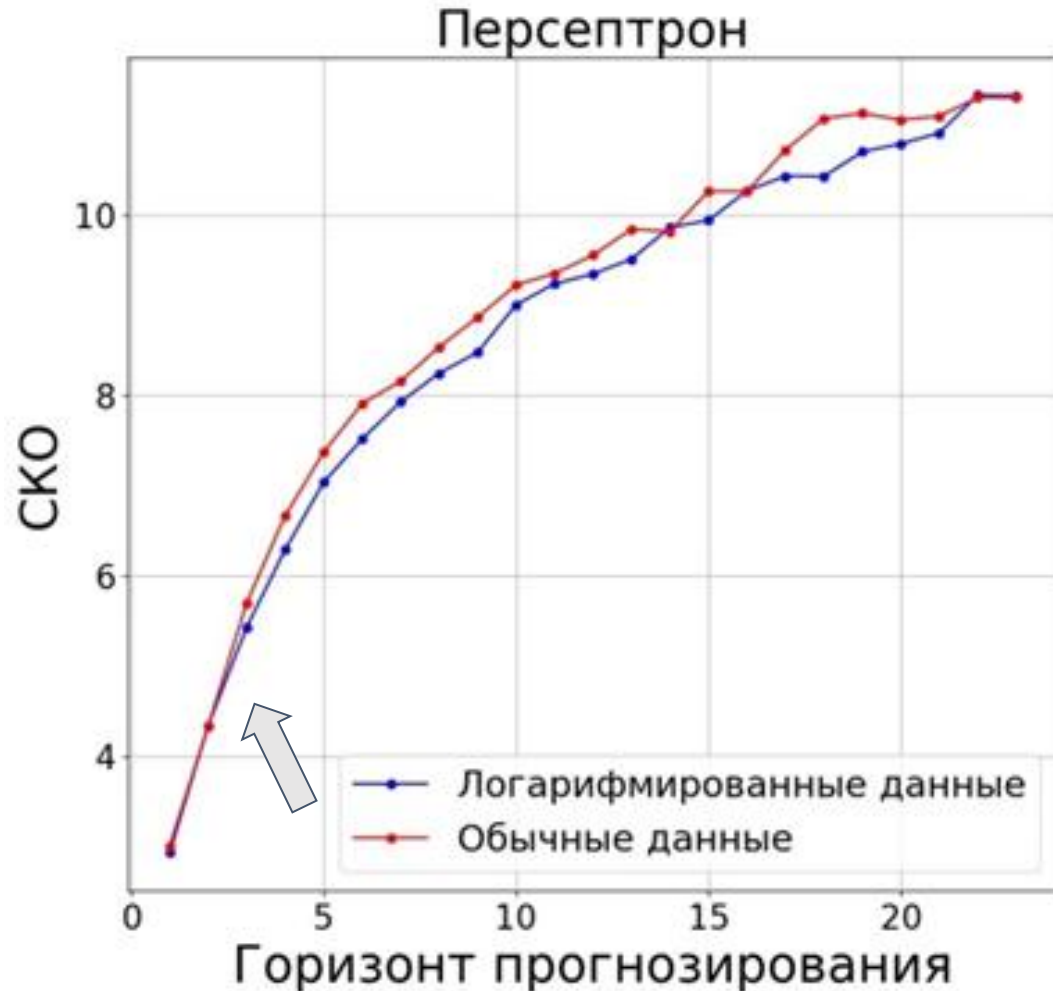
Индекс Dst в исходном представлении



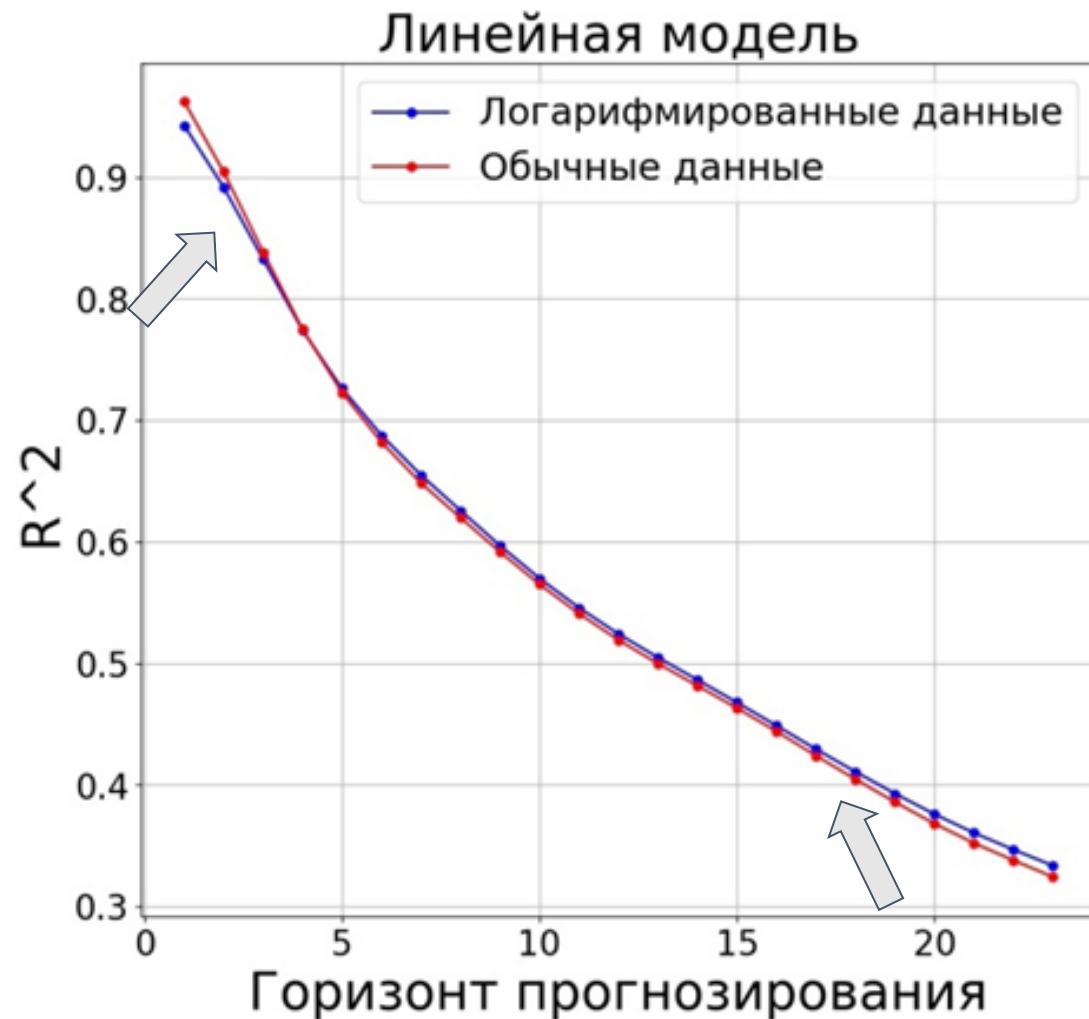
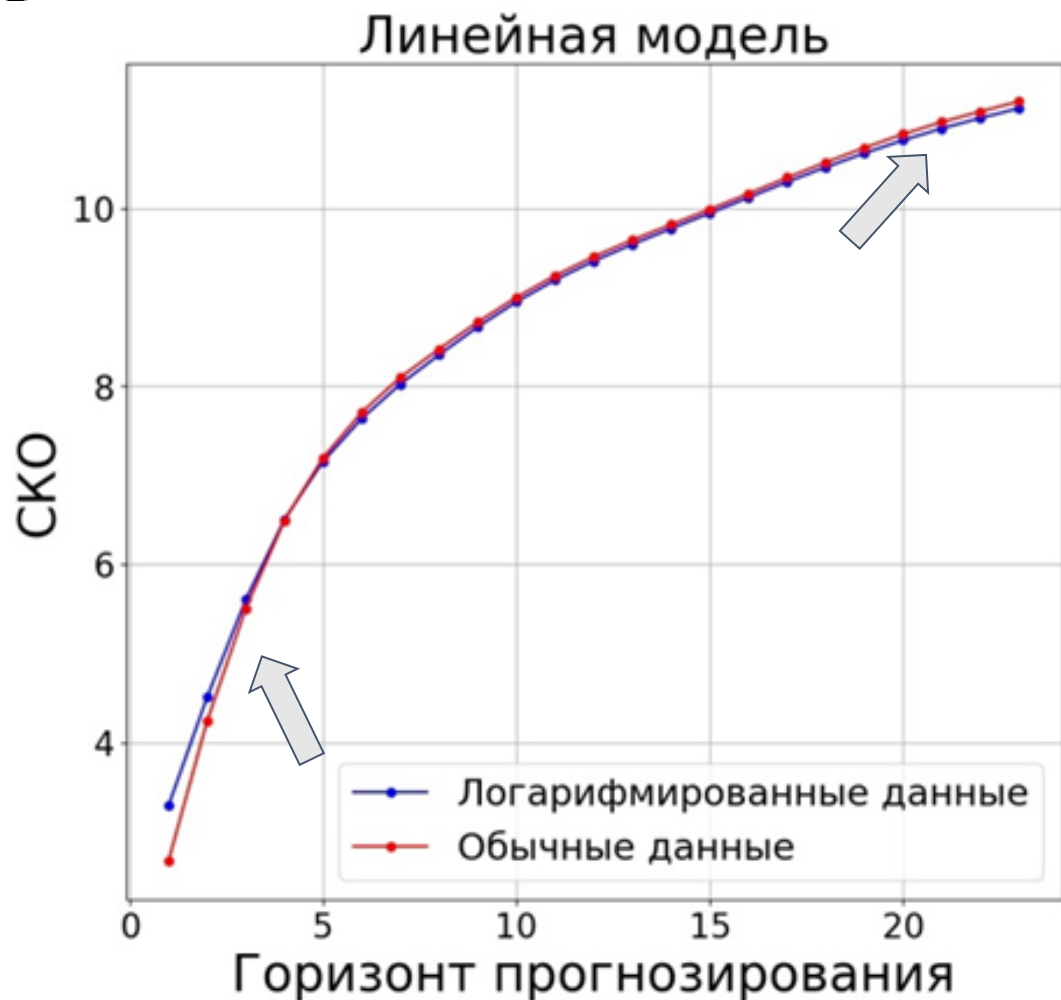
Зависимость качества прогнозирования от горизонта прогноза. Индекс Dst в логарифмическом представлении.



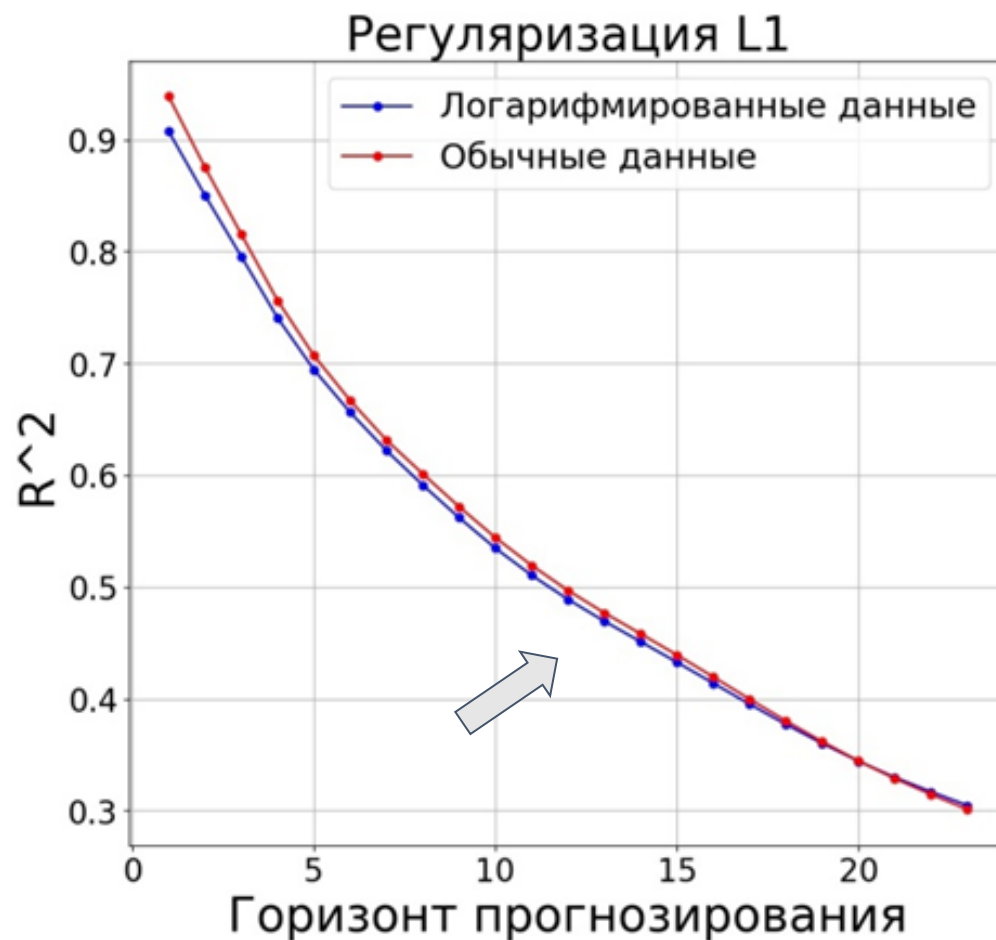
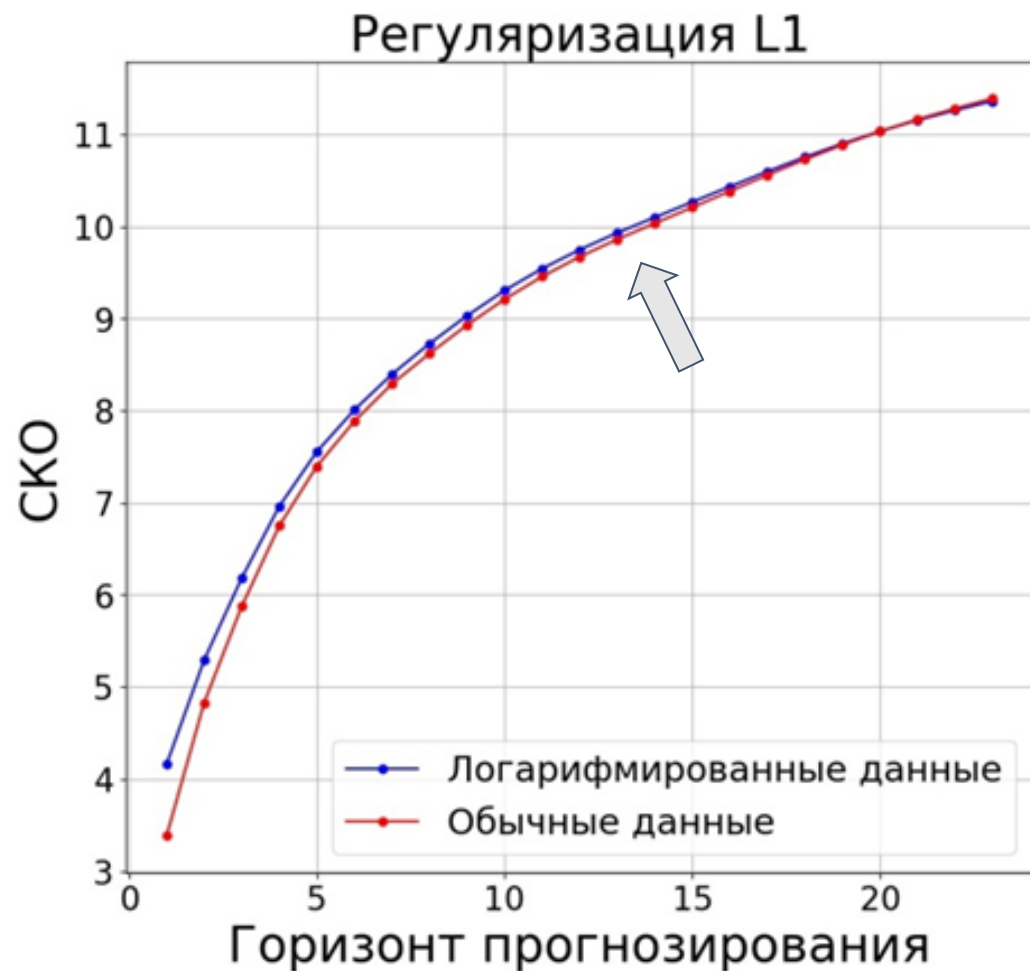
Сравнение зависимостей качества прогнозирования от горизонта прогноза для исходного и логарифмированного представления индекса Dst.



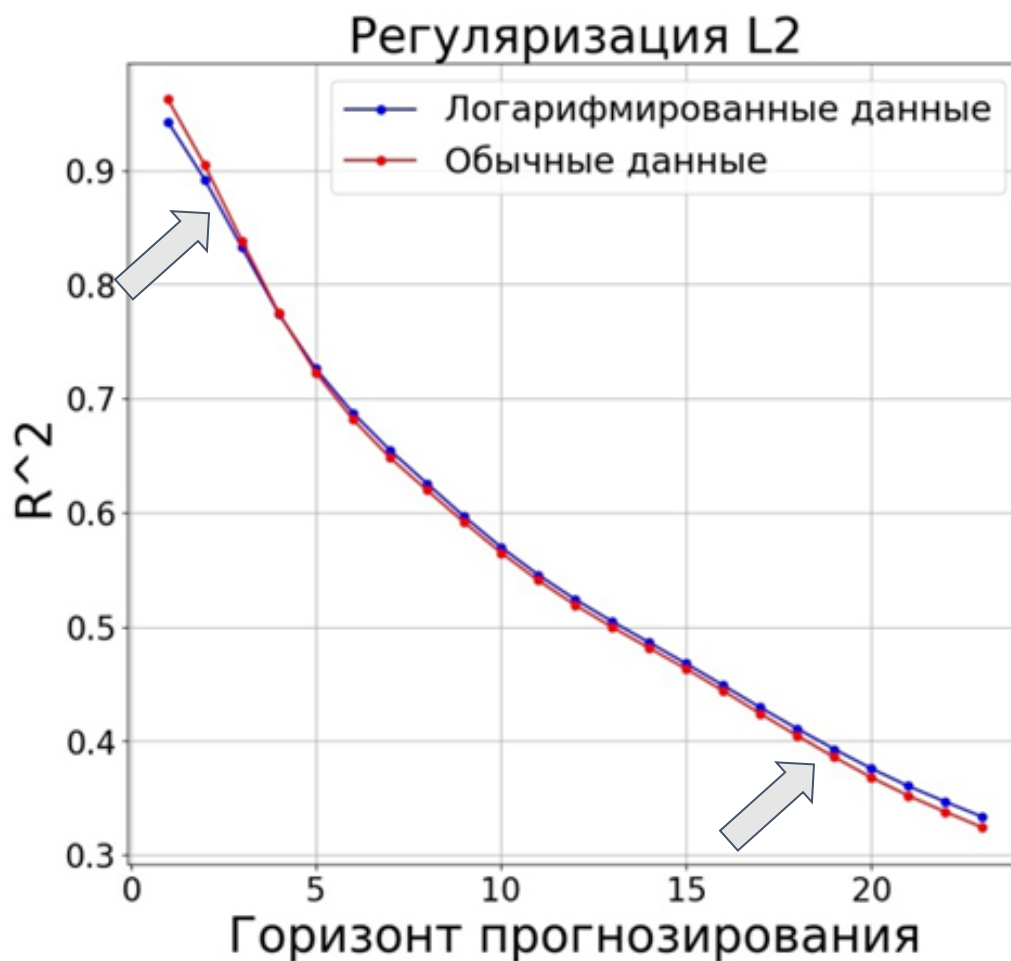
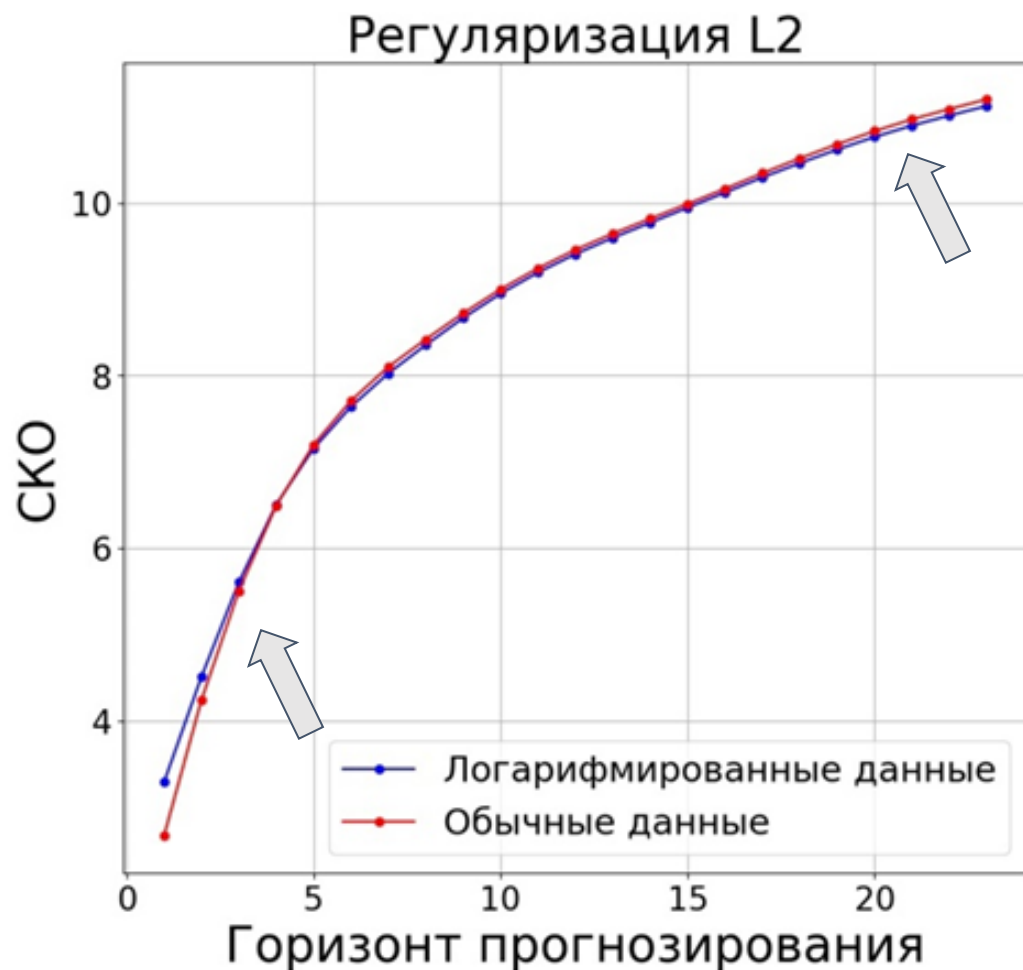
Сравнение зависимостей качества прогнозирования от горизонта прогноза для исходного и логарифмированного представления индекса Dst.



Сравнение зависимостей качества прогнозирования от горизонта прогноза для исходного и логарифмированного представления индекса Dst.



Сравнение зависимостей качества прогнозирования от горизонта прогноза для исходного и логарифмированного представления индекса Dst.



Выводы

По результатам настоящей курсовой работы могут быть сделаны следующие выводы:

- 1) Для всех построенных моделей качество прогнозирования монотонно убывает с увеличением горизонта прогноза. При этом минимально приемлемым можно считать качество прогнозирования при горизонте прогноза не более 12 часов.
- 2) Построенные модели могут быть ранжированы по качеству прогнозирования в следующем порядке: многослойный персептрон, гребневая регрессия, линейная регрессия без регуляризации, лассо регрессия, тривиальная модель.



Выводы

3) Использование логарифмированных данных повышает точность прогнозирования при горизонте прогноза 2,5 часов и более. При меньших величинах горизонта качество прогнозирования ухудшается на линейных моделях и почти не меняется на персептроне, что может быть связано с низкой нелинейностью требуемого отображения, которое более эффективно аппроксимируется в исходном не логарифмированном представлении.



Спасибо за внимание

