**SVM-подход к решению задачи восстановления регрессии** (**SVR**)

План занятия

1. *Постановка задачи восстановления регрессии.*

***х*** ∈ *Rn*: *P*(***x***) . Вектор ***х***– вектор контролируемых факторов.

*у* ∈ *R* :  *P*(*y*|***x***). Число *у* - отклик.

Существует функция регрессии

*y*\*(***x***) = ∫ y P(*y*|***x***) *dy*.

Требуется по случайной выборке *l* независимых пар (***x****1*, *y1*), … , (***x****l*, *yl*), таких, что

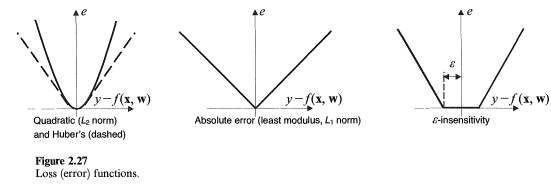
*yi* = *y\**(***x****i*) + ξ*i*, где ξ*i*,  *i*=1, …, *l* – случайные ошибки (шум),

восстановить регрессию, т.е. в заданном классе функций {*f*(***x***,α} отыскать функцию *f*(***x***,α*\**), наиболее близкую к регрессии *y\**(***x***).

2. *Различные функции потерь и соответствующие плотности распределения шума.*

Выбор функции потерь, определяющей, каким образом штрафуются ошибки обучения, должен быть содержательно обоснован для каждой конкретной задачи прогнозирования. Поскольку различным функциям потерь соответствуют различные методы обучения, целесообразно избегать сложных функций, которые могут привести к трудной оптимизационной задаче.

Обычно используют симметричные функции потерь, такие как *ε* - нечувствительная, юберовская, абсолютная и квадратичная ошибки.

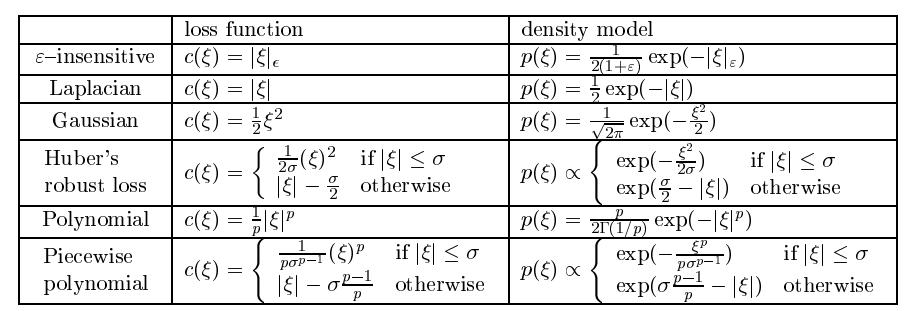


Вид функции потерь связан с априорными предположениями о распределении шума в регрессионной модели. (См.[1] стр.10-17 и/или.[2] стр.204-206)

А именно, оптимальная согласно принципу максимального правдоподобия функция потерь имеет вид:

*L*(*f*(***x***),*y*) = –ln *p*(*y*–*f*(***x***)), где *p*(*y*–*f*(***x***)) ≡ *p*(ξ) – плотность распределения шума.





3. ***SVM****-подход к решению задачи восстановления регрессии с-нечувствительной функцией потерь.*

Пусть (***x****1*, *y1*), … , (***x****l*, *yl*) ∈ *Rn*×*R* − тренировочная последовательность (ТП). Требуется найти наиболее гладкую функцию *f*(***x***), величина отклонения значений которой от наблюдаемого *y* на всей тренировочной последовательности не превышает заданное число *ε*.

Будем искать линейную оценку функции регрессии в виде: *f*(***x***)=(***w***, ***x***)+*b*, ***w*** ∈*Rn* , *b*∈*R*.Мерой гладкости класса линейных функций может служить евклидова норма ||***w***||.

Для ослабления жестких ограничений введем переменные ξ*i*, ξ*i\** ≥ 0. Тогда, согласно принципу структурной минимизации риска, задача восстановления регрессии сводится к задаче оптимизации:

 (1)

Параметр регуляризации *С* > 0устанавливает баланс между допустимой величиной ошибки и сложностью класса функций, из которого выбирается решение *f*(***x***). Формулировка (1) соответствует использованию так называемой ε-нечувствительной функции потерь |*f*(***x****i*)–*yi*|ε , где

.

(*ε*-insensitive loss function)

Однофакторная линейная регрессия и ε-нечувствительная функция потерь представлены на рисунке:

C:\Users\Serge\Desktop\Рис.1.tif

Для задачи выпуклого программирования (1) лагранжиан имеет вид:

 (2)

α*i*(\*)0, η*i*(\*)0, ξ*i*(\*)0, где ζ(\*) обозначает как переменную ζ, так и ζ\*.

Вычисляя частные производные лагранжиана по ***w***, *b*, ξ*i*(\*)и приравнивая их нулю, получим:

 (3)

 (4)  (5)

Подставляя (3)-(5) в (2), приходим к двойственной задаче:

 (6)

Условия оптимальности Каруша-Куна-Таккера (ККТ)в этом случае будут необходимыми и достаточными. Наряду с соотношениями (3)–(5), условия ККТ включают условия комплементарности:

 (7)

 (8)

Понятно, что α*i*α*i*\*=0 для всех *i* .

Из условий ККТ следует:

только образцы, для которых α*i*(\*) = *C*, лежат вне ε-области вокруг *f*(***x***);

для α*i*(\*)∈(0, *C*) соответствующие ξ*i*(\*)=0, при этом легко можно вычислить сдвиг *b*;

для всех образцов, лежащих внутри ε-области (ε-tube), α*i*(\*)=0, что свидетельствует о разреженности SV-решения.

Таким образом, вSV-разложение войдет только часть векторов тренировочной последовательности. Они и являются **опорными векторами**.

Условия ККТ позволяют найти вектор ***w*** с использованием только тренировочной последовательности:

 (9)

и, следовательно, решение задачи восстановления линейной регрессии имеет вид:

 (10)

Сложность представления аппроксимирующей функции (10) не зависит от размерности входных векторов, а зависит только от числа опорных векторов.

4. ***SVM****-подход к решению задачи восстановления регрессии с квадратичной функцией потерь (****LS\_SVR****).*

Использование квадратичной функции потерь приводит к задаче оптимизации вида:



,

что соответствует гребневой регрессии, а минимизация эмпирического риска – методу наименьших квадратов. Параметр регуляризации γ>0 связан с введенным ранее параметром *C* соотношением γ = (2*С*)/*l*.

Составим функцию Лагранжа:

,

где **α**∈*R* – вектор множителей Лагранжа.

Выпишем условия оптимальности для *k*=1,...,*n*, *i*=1,...,*l*:

 (11)

 (12)

 (13)

 (14)

После исключения из соотношений (14) ξ*i*, *i*=1, …, *l*, и ***w*** согласно (13) и (11) получаем систему линейных алгебраических уравнений (СЛАУ) порядка (*l*+1) с вектором неизвестных (**α*,*** *b*):

,

где ***e*** – единичный вектор-столбец размерности *l*, *E* - единичная матрица *l*×*l*, *K* – ядерная матрица *l*×*l* (ядро линейное), ***y*** – вектор-столбец наблюдаемых откликов.

Решение задачи восстановления линейной регрессии по-прежнему будет иметь вид:

.

Однако SVM-решение, построенное на основе квадратичной функции потерь, в общем случае абсолютно плотно: α*i*≠0 для всех *i*. Т.е. в данном случае свойство разреженности решения не имеет места. Таким образом реализация метода требует настолько больших временных затрат и такого объема памяти, что метод становится непригодным для практического использования. На практике применяют различные "экономичные" алгоритмы для получения разреженной аппроксимации полного ядерного разложения.

5. *Модификация стандартного* ***SVM****-подхода к решению задачи восстановления регрессии: -****SVR****. Интерпретация параметра .*

Одной из полезных модификаций стандартной SVM-задачи является так называемая ν *-*SVM.

В случае задачи восстановления регрессии с ε-нечувствительной функцией потерь такая модификация функционала регуляризованного риска, при которой ε становится переменной, подлежащей оптимизации, дает:

 (15)



Значение параметра ν∈[0;1] является нижней границей доли входных векторов, которые становятся опорными, и верхней границей доли ошибок.

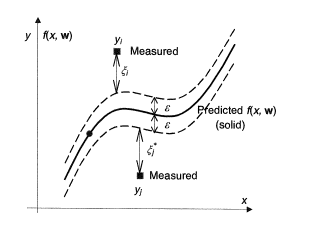
Преимущество ν*-*SVM-подхода по сравнению со стандартным состоит в следующем: значение ε **настраивается автоматически**; ν ∈ [0;1]; параметр ν имеет четкую интерпретацию.

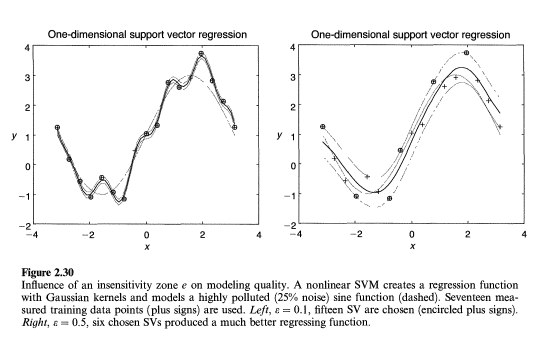
6. *Нелинейная* ***SVR***.

И в процессе обучения, и при прогнозировании в рассмотренных задачах входные объекты появляются только в виде скалярных произведений. Векторы тренировочной последовательности нелинейно отображаются в пространство признаков сколь угодно большой размерности. Для гильбертовых пространств с репродуктивным ядром существует высокоэффективный прием вычисления скалярных произведений, который состоит в использовании ядерных функций. Поэтому любой линейный алгоритм, использующий только скалярные произведения, может быть неявно выполнен в пространстве признаков, надо лишь каждое скалярное произведение заменить нелинейным ядром.

Процесс обучения для нелинейной задачи (как классификации, так и восстановления регрессии) аналогичен процессу решения соответствующей линейной задачи, но скалярные произведения (***x***, ***x***′) заменяются значениями ядра *k*(***x***, ***x***′), и вектор ***w*** теперь принадлежит пространству признаков.

Однофакторная нелинейная регрессия с ε-нечувствительной функцией потерь представлена на рисунке:





Истинная зависимость (бледная прерывистая линия), ее прогноз (сплошная жирная кривая) и *ε*-зона. Тренировочная последовательность (+). Опорные векторы(.

Рисунок иллюстрирует, как увеличение ε-области сглаживает влияние шума (данные сильно зашумлены) на модель (прогноз истинной закономерности, синусоиды в данном случае). Увеличение ε означает ослабление требований к точности аппроксимации на ТП. Это ведёт также к уменьшению числа опорных векторов. Слева: ε = 0.1, выбраны 15 опорных векторов. Справа: ε = 0.5, выбраны 6 опорных векторов, которые обеспечивают лучший прогноз.

На рисунке показаны построенные по точным данным машины опорных векторов для различных значений параметра ε: прогноз истинной закономерности (сплошная линия) по данным тренировочной последовательности (мелкие точки), а также образцы (жирные точки), определяющие опорные векторы. Заметим, что с уменьшением ε растет число опорных векторов. Вверху: слева ε = 0.5, справа ε = 0.2; внизу: слева ε = 0.1, справа ε = 0.02.



Литература

1. *A Tutorial on Support Vector Regression*. Alexander Smola and Bernhard Scholkopf. NeuroCOLT2 Technical Report Series. October,1998.

2. *A tutorial on support vector regression*. A . J. Smola and B. Scholkopf Statistics and Computing 14: 199–222, 2004.