

npregress intro - ノンパラメトリック回帰の概要 【評価版】

ノンパラメトリック回帰を使うと、共変量が与えられたときのアウトカム変数の平均値がどのような関数形に従うかについて不確かな場合でも、そのモデル化を可能にします。ノンパラメトリックなカーネルベースの推定法はバイアスと分散をトレードオフさせるような最適なバンド幅パラメータに依存しています。ノンパラメトリック回帰は関数形の指定に関する誤りに対して頑健な推定手法を提供するものですが、より多くの観測データと計算時間を必要とするという代償を伴うものでもあります。

npregress は局所的線形 (local linear) と局所的定数 (local constant) という 2 種類のノンパラメトリック推定法を実装しています。本 whitepaper ではそれらの概要を紹介します。

1. ノンパラメトリック回帰

1. ノンパラメトリック回帰

ノンパラメトリック回帰は共変量が与えられたときのアウトカム変数の平均値がどのような関数形に従うかについて不確かな場合に使用されます。例えば線形回帰を用いて推定を行う場合、アウトカム変数の平均値の関数形としては共変量の線形結合であること、あるいは共変量とそれらの交互作用の線形結合であることが仮定されます。パラメトリック回帰にしろノンパラメトリック回帰にしろ、それらは共変量の種々の値に対する平均値を推定します。今、図 1 に示されるような擬似的なデータについて考えてみましょう。 x のすべての値に対するアウトカム変数の平均値を表す曲線が散布図上に示されています。



図 1

図 1 に示されたデータの平均値は x に関し線形ではないため、線形回帰を行ったとするなら、それはアウトカム変数に対する共変量 x の効果を正確にグラフ化したものとは言えなくなります。実際、このデータに対し線形回帰をフィットさせると図 2 のようなプロットがもたらされることになります。

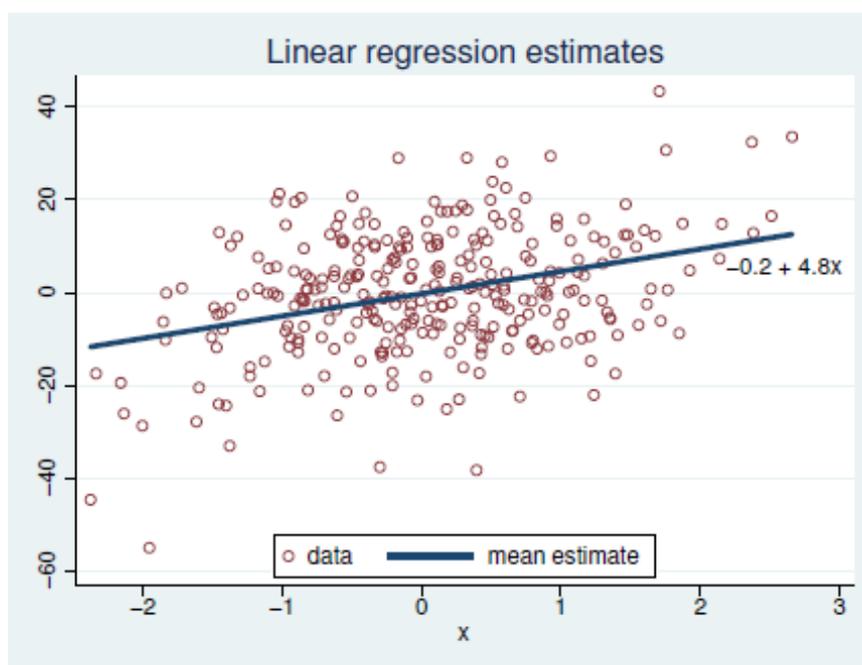


図 2

x が変化したときの応答変数予測値の変化は正の値で一定であるわけですが、真の平均値は非線形な形で変化しています。平均値の関数形に関する仮定が正しくなかった場合、得られる推定値は不整合なものとなります。これに対し `npregress` を用いてモデルをフィットさせ推定結果をグラフ化すると図 3 のようなプロットが得られます。

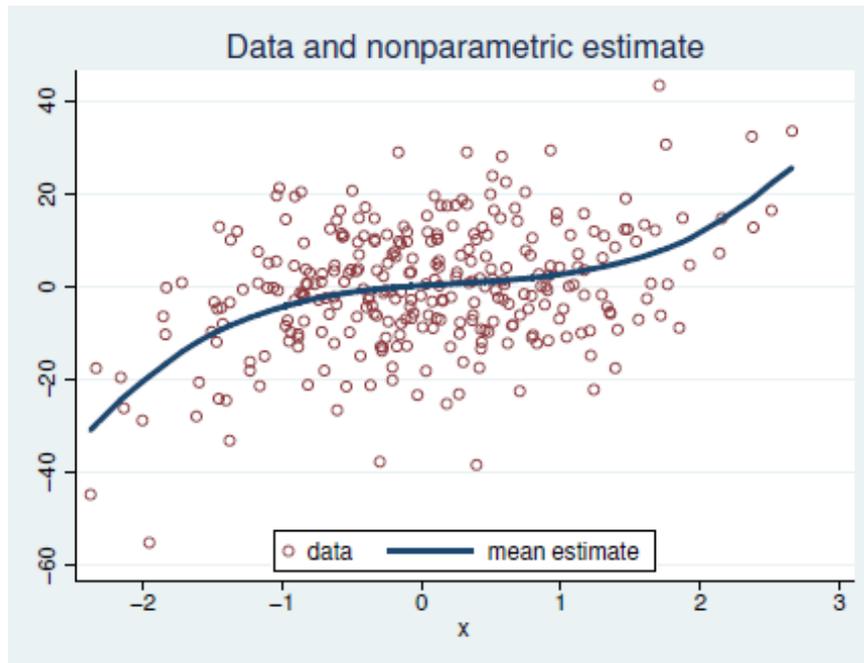


図 3

`npregress` はアウトカム変数と共変量間の関係を正しく表現できています。ノンパラメトリック回帰による推定結果は真の関数が十分滑らかなものである限り整合性が取れたものとなります。線形回帰の前提が正しかった場合であってもノンパラメトリック回帰は整合性のある結果をもたらしますが、効率はより低いものとなります。

ノンパラメトリック回帰は関数形の指定に関する誤りに対して頑健な推定手法を提供するものですが、より多くの観測データと計算時間を必要とするという代償を伴うものでもあります。この代償コストは共変量の数と共に増加します — 次元の呪い (curse of dimensionality) —。

以下においては図 3 に示されるような推定結果をどのように求めたかについて説明して行きます。具体的には種々の x の値において平均値を計算することにより関数の形状を浮き彫りにして行くわけです。

評価版では割愛しています。

■