ノンパラメトリックモデルは, 実際には実装できない単なる理論的抽象化 (考え得る全ての確率分布を検索するアルゴリズムなど) である場合がある. ただし, ノンパラメトリックモデルの複雑さをトレーニングセットサイズの関数にすることで, 実用的なノンパラメトリックモデルを設計することもできる. そのようなアルゴリズムの一例は, 最近傍回帰である. 重みの固定長ベクトルを持つ線形回帰とは異なり, 最近傍回帰モデルはトレーニングセットから $\boldsymbol{X}$ と $\boldsymbol{y}$ を単に保持する. テストポイント $\boldsymbol{x}$ を分類するよう求められると, モデルはトレーニングセット内の最も近いエントリを検索し, 関連する回帰ターゲットを返す. 言い換えれば, $\hat{y} = y_i$　ここで, $i = argmin||\boldsymbol{X}_{i,:} - \boldsymbol{x}||_2^2$ である. (2015 年, Goldberger らによれば) このアルゴリズムは学習された距離メトリックのような L^2 ノルム以外の距離メトリックにも一般化できる. アルゴリズムが, 最も近い全ての点 $\boldsymbol{X}_{i,:}$ の y_i 値を平均することで同点を解読できる場合, このアルゴリズムは任意の回帰データセットに対してトレーニング誤差 (同一の 2 つの入力が, 異なる出力に関連づけられている場合, この誤差は 0 より大きくなる可能性がある) を最小限に抑えることができる.

最後に, パラメトリック学習アルゴリズムを包含する, 必要に応じてパラメータ数を増やす別のアルゴリズムを作ることで, ノンパラメトリック学習アルゴリズムを作成することもできる. 例えば, 入力の多項式展開に加えて, 線形回帰によって学習された多項式の次数を変更する学習の外側のループが想定される.

理想的なモデルは, データを生成する真の確率分布を単に知るオラクルである. このようなモデルであっても, 分布には依然としてノイズが存在する可能性があるため, 多くの問題で何らかの誤差が生じる可能性がある. 教師あり学習の場合, $\boldsymbol{x}$ から y へのマッピングは本質的に確率的であるか, y が $\boldsymbol{x}$ に含まれる変数以外の他の変数を含む決定論的関数である可能性がある. 真の分布 $p(\boldsymbol{x}, y)$ から予測を行うオラクルによって発生する誤差はベイズ誤差と呼ばれる.

トレーニングセットのサイズが変化すると, トレーニングと汎化の誤差も変化する. トレーニングサンプルの数が増加しても予期される汎化誤差が増加することはない. ノンパラメトリックモデルの場合, 可能な限り最良の誤差が得られるまで, より多くのデータがより適切な一般化をもたらす. 最適な容量に満たない固定パラメトリックモデルは, ベイズ誤差を超える誤差の値に漸近する. 図 5.4 を参照ください. モデルが最適な容量を持っていたとしても, トレーニング誤差と汎化誤差の間に大きなギャップがあることに留意してください. この状況では, より多くのトレーニングサンプルを収集することで, このギャップを縮めることができるかもしれない.

5.2.1 ノーフリーランチ定理

学習理論では, 機械学習アルゴリズムは有限のサンプルのトレーニングセットから適切に一般化できると主張する. これは, 論理のいくつかの基本原則に矛盾しているように思える. 帰納的推論, つまり限られたサンプルのセットから一般規則を推測することは論理的に妥当ではない. セットの全てのメンバを記述するルールを論理的に推論するには, そのセットの全てのメンバに関数情報が必要である.

純粋に論理的な推論で使用される完全に特定のルールではなく, 確率的なルールのみを提供することで, 機械学習は部分的にこの問題を回避する. 機械学習は, 対象となるセットのほとんどのメンバについて正しいであろうルールを見つけることが期待される.

残念ながら, これでも問題全体が解決されるわけではない. (1996 年, Wolpert によると) 機械学習のノーフリーランチ定理では, 考え得る全てのデータの生成分布を平均すると, 前もって観察されていない点を分類する際に, 全ての分類アルゴリズムの誤り率が等しくなると述べられている. 言い換えれば, ある意味で, 普遍的に他のアルゴリズムより優れた機械学習アルゴリズムは存在しないということである. 我々が思いつく最も洗練されたアルゴリズムは, 全ての点が同じクラスに属することを単に予測するのと同じく (考え得る全てのタスクにわたって) 同様の平均性能を備えている.

図 5.4: トレーニングデータセットのサイズがトレーニングとテスト誤差, 適切なモデルの容量に及ぼす影響. 5 次多項式に適度な量のノイズを加えることで合成回帰問題を構築し, 単一のテストセットを生成し, 次にいくつかの異なるサイズのトレーニングセットを生成した. 95%の信頼空間を示す誤差バーを描画するために, サ

イズごとに 40 の異なるトレーニングセットを生成した. (上) トレーニング上の MSE と 2 つの異なるモデルのテストセット: 二次モデルと, テスト誤差を最小限に抑えるように選択された字数を持つモデル. どちらも閉形式で適合する. 二次モデルの場合, トレーニングセットのサイズが増加するにつれて, トレーニング誤差も増加する. これは, データセットが大規模であるほど適合するのが困難になるためである. 同時に, トレーニングデータと一致する不正確な仮説が少なくなるため, テスト誤差は減少する. 二次モデルにはタスクを解決するのに十分な能力がないため, そのテスト誤差は高い値に漸近する. 最適な容量でのテスト誤差は, ベイズ誤差に漸近する. トレーニングアルゴリズムがトレーニングデータの特定のインスタンスを記憶する能力があるため, トレーニング誤差がベイズ誤差を下回る可能性がある. トレーニングサイズが無限に増加すると, 任意の固定容量のモデル (ここでは, 二次モデル) のトレーニング誤差は少なくともベイズ誤差まで上昇しなければならない. トレーニングセットのサイズが増加するにつれて, (下) 最適な容量 (ここでは, 最適な多項式回帰子の次数として示される) が増加する. タスクを解決するのに十分な複雑さに達すると最適な容量は頭打ちになる.

幸いなことに, 全ての可能なデータの生成分布を平均した場合にのみ当てはまる. 現実世界のアプリケーションで遭遇するような確率分布の種類について仮説を立てると, これらの分布で適切に実行される学習アルゴリズムを設計できる.

これは, 機械学習研究の目標が, 普遍的な学習アルゴリズムや絶対的に最適な学習アルゴリズムを求めることではないことを意味する. 代わりに, 我々の目標は AI エージェントが経験する「現実世界」にどのような種類の分布が関連しているのか, また, 我々が関心を持つデータの生成分布から抽出されたデータに対してどのような種類の機械学習アルゴリズムが適切に機能するかを理解することである.

5.2.2 正則化

ノーフリーランチ定理は, 特定のタスクで適切に機能するように機械学習アルゴリズムを設計する必要があることを意味している. これは, 学習アルゴリズムに一連の優先順位を組み込むことで実現される. これらの優先順位がアルゴリズムに解決を求める学習問題と一致している場合, アルゴリズムの性能は向上する.

これまで説明してきた学習アルゴリズムを修正する唯一の方法は, 学習アルゴリズムが選択可能な解の仮説空間に関数を追加あるいは削除することで, モデルの容量を増減することである. 回帰問題の多項式の次数を増減させる具体的な例を挙げました. これまで説明してきた見解は過度に単純化し過ぎている.

我々のアルゴリズムの動作は, 仮設空間で許可される関数の集合をどれだけ大きくするかだけでなく, それらの関数の具体的な性質によっても強く影響される. これまで研究してきた学習アルゴリズムである線形回帰には, 入力の線形関数の集合で構成される仮設空間がある. これらの線形関数は, 入力と出力の間の関係が実際に線形に近い問題に対して非常に役立つ. 非常に非線形的な動作をする問題にはあまり役に立たない. 例えば, 線形回帰を使用して x から $sin(x)$ を予測しようとしてもあまりうまく機能しない. したがって, 解を引き出すことができる関数の種類を選択し, これらの関数の量を制御することで, アルゴリズムの性能を制御できる.