誤差関数について

ここで言う誤差とは、厳密に言うと、分布と分布の差を測ると言うこと。

今、との分布があるとする。

との分布の差を測るために、次の式が用いられている。

(1)が、の時、０以上であることを示したい。

ここで、イェンセンの不等式を思い出そう。

|  |
| --- |
| を実数上の凸関数とする。  離散の場合  を=1を満たす正の実数列とする。また、を実数の列とする。このとき以下が成り立つ。  連続な場合  をを満たす実数上の可積分関数とする。また、を実数上の可積分関数とする。そのとき次が成り立つ。  このとき、を確率密度関数とみると、イェンセンの不等式は次のようになる。  参考URL：<https://ja.wikipedia.org/wiki/イェンセンの不等式> |

イェンセンの不等式を用いると、（1）は次のように変形できる。

よって示せた。

今、2つの確率分布を、自分が作る確率モデルと神の確率分布として、が何かを考えてみよう。

ここで、サンプル近似を思い出そう。

|  |
| --- |
| X〜（確率分布Xが確率分布に従う、母集団からXをサンプリングした）とする。  この時、における期待値を近似できる。  ただし、は、母集団からサンプリングしたものとする。  その理由は、母集団でのデータ（全てのデータ）の平均の方が、期待値（平均は厳密であるから） |

＝+

ここで、について考えると、は神の確率モデルだったので、最大化、最小化はできない。つまり、制御下にない。

について考えると、は、自分が作る確率モデルだったで、最大化、最小化が可能である。つまり制御可能。

回帰問題では、神の確率モデルとの差を小さくしたい。

つまり、をできるだけ最小にするパラメータを求めたい。より厳密の言うと、が最小になるパラメータを求めたい。

を最小にしたいと言うことは、を最大にすることと同値である。

ここで、をサンプル近似すると、次のようになる。

(2)の式の表す重要なところは、神から与えられたデータに基づいて考えられると言うことである。

また、(2)の式は、で与えられており、尤度最大化（最適化）が神の確率分布との差を最小にすることと同値であることを表している。