

Homework 1

余磊 SA21011091

1.2

《撰写手册》要求数学符号遵循 GB/T 3102.11—1993《物理科学和技术中使用的数学符号》。该标准参照采纳 ISO 31-11:1992, 但是与 T_EX 默认的美国数学学会 (AMS) 的符号习惯有所区别。具体地来说主要有以下差异:

If the integral of function f is measurable and non-negative, we define its (extended) Lebesgue integral by

$$\int f = \sup_g \int g,$$

where the supremum is taken over all measurable functions g such that $0 \leq g \leq f$, and where g is bounded and supported on a set of finite measure. Suppose $\{f_n\}$ is a sequence of measurable functions with $f_n \geq 0$. If $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for a.e. x , then

$$\int f \leq \liminf_{n \rightarrow \infty} \int f_n.$$

Suppose f is a non-negative measurable function, and $\{f_n\}$ a sequence of non-negative measurable functions with $f_n(x) \leq f(x)$ and $f_n(x) \rightarrow f(x)$ for almost every x . Then

$$\lim_{n \rightarrow \infty} \int f_n = \int f.$$

Suppose f is integrable on \mathbb{R}^d . Then for every $\epsilon > 0$:

i. There exists a set of finite measure B (a ball, for example) such that

$$\int_{B^c} |f| < \epsilon.$$

ii. There is a $\delta > 0$ such that

$$\int_E |f| < \epsilon \quad \text{whenever } m(E) < \delta.$$

Suppose $\{f_n\}$ is a sequence of measurable functions such that $f_n(x) \rightarrow f(x)$ a.e. x , as n tends to infinity. If $|f_n(x)| \leq g(x)$, where g is integrable, then

$$\int |f_n - f| \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and consequently

$$\int f_n \rightarrow \int f \quad \text{as } n \rightarrow \infty.$$

1.3

对样本空间放宽约束，将满足样本数量最多的假设视为最优假设。

2.1

训练集各有 70% 的正例和反例，共有 $C_{500}^{350} * C_{500}^{350}$ 种。

2.6

ROC 曲线上的任意一点 (FPP, TPP) 对应的错误率是：

$$\begin{aligned} \text{Error rate} &= \frac{FP + FN}{m^+ + m^-} \\ &= \frac{m^- \cdot FPP + m^+ \cdot (1 - TPP)}{m^+ + m^-} \end{aligned}$$

故 ROC 曲线上越接近 (0,1) 的对应的错误率越小，越接近 (1,0) 的点对应的错误率越大。

2.8

min-max 规范化

优点: 简单; 加入新值后无需全部重新计算

缺点: 对离群点敏感

z-score 规范化

优点: 对离群点不敏感

缺点: 复杂; 加入新值后需全部重新计算