MATH524 – Spring 2025

Problem Set: Week 10

1. (**Conditional expectation**) Show that $m(x) = \mathbb{E}[Y|X = x]$ minimizes

$$\mathbb{E}[|f(X) - Y|^2]$$

over all measurable functions $f : \mathbb{R}^d \to \mathbb{R}$. That is, the regression function minimizes the expected squared-loss.

2. (**Error decomposition**) Let $\mathcal{F}_n$ be some class of functions $f : \mathbb{R}^d \to \mathbb{R}$ that depend in some way on the data $\{(X_i, Y_i)\}_{i=1}^n$ that is generated by a standard regression model:

$$Y_i = m(X_i) + \varepsilon_i.$$

Consider the regression estimator m_n that satisfies

$$m_n = \operatorname*{arg\,min}_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^n |f(X_i) - Y_i|^2.$$

Show the following

$$\int |m_n(x) - m(x)|^2 \mu(\,\mathrm{d}x) \leq 2 \sup_{f \in \mathcal{F}_n} \left| \frac{1}{n} \sum_{i=1}^n |f(X_i) - Y_i|^2 - \mathbb{E}[(f(X) - Y)^2] \right| + \inf_{f \in \mathcal{F}_n} \int |f(x) - m(x)|^2 \mu(\,\mathrm{d}x).$$