Mathematics for New Technologies in Finance

Solution sheet 6

Exercise 6.1 (Bayesian optimization)

(a) Recall the definition of prior, likelihood, posterior, and evidence distributions in bayesian statistics.

(b) Consider linear model on $\mathbb{R}: Y \sim \theta X + Z, \theta \sim \mathcal{N}(0,1), Z \sim \mathcal{N}(0,1)$ and θ independent with X. Compute $p_\theta(y \mid x)$ and $p(\theta \mid x, y)$. Prove that maximizing the posterior $p(\theta \mid x, y)$ is exactly doing Ridge regression (fix λ here).

(c) Consider Lasso regression, what is the prior under Bayesian perspective? Please calculate the posterior under this prior.

(d) Would you expect a sparser weight or denser weight using Lasso regression instead of Ridge regression.

Solution 6.1

(a) Posterior = Likelihood * Prior / Evidence

(b) See the proof here.

(c) Suppose we have data points $y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \epsilon_i$. where $\epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$. The likelihood for the data is

$$\mathcal{L}(Y \mid X, \beta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\epsilon_i^2}{2\sigma^2}\right) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n \epsilon_i^2\right) \tag{1}$$

The prior for β follows Laplace distribution(also known as double-exponential distribution) with a zero mean and common scale parameter $b : p(\beta) = (1/2b)\exp(-|\beta|/b)$. Multiplying the prior distribution with the likelihood we get the posterior distribution

$$\mathcal{L}(Y \mid X, \beta) p(\beta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n \epsilon_i^2\right)\left[\frac{1}{2b}\exp\left(-\frac{|\beta|}{b}\right)\right] \tag{2}$$

(d) Sparser for Lasso.

Exercise 6.2 (Implied volatility) The Black-Scholes formula provides a relationship between the price of a European Call option $C(K,T)$ and volatility $\sigma(K,T)$ for fixed price of underlying S_0, strike K, and maturity T. It is an important transformation in Finance to calculate from $C(K,T)$ the *implied volatility* $\sigma(K,T)$. Proceed in the following steps:

- Define a Gamma prior on implied volatility.
- Define a likelihood, which predicts the price given an implied volatility.
- Construct a posterior via Bayes formula and sample from it via Langevin dynamics. Interpret the resulting algorithm from the perspective of stochastic gradient descent.

Solution 6.2

- Assume $\pi(\sigma) = \frac{\sigma^{\alpha-1} exp(-\beta\sigma)\beta^{\alpha}}{\Gamma(\alpha)}, \delta > 0$
- Denote the call price calculated using Black-Scholes formula based on the implied volatility(σ) as $C(\sigma)$. Then if the market price is assumed to follow a log-normal distribution, we have the likelihood:

$$\mathcal{L}(\sigma \mid C(K,T)) = \frac{1}{\sigma\sqrt{2\pi}} exp(\frac{-(ln(C(K,T) - ln(C(\sigma)))^2}{2\sigma^2}) \tag{3}$$

- Using Bayes formula, we have the posterior distribution

$$\pi(\sigma \mid C(K,T)) \propto \mathcal{L}(\sigma \mid C(K,T))\pi(\sigma) \tag{4}$$

- Define learning rate η and noise term $\eta^{1/2}\varepsilon_t$, iterative update implied volatility

$$\sigma_{t+1} = \sigma_t - \eta\nabla(-log(\pi(\sigma \mid C(K,T))) + \eta^{1/2}\varepsilon_t \tag{5}$$

This algorithm leverages Bayesian statistics to estimate the implied volatility of a call option by incorporating prior knowledge (through the Gamma prior) and the observed market price (through the likelihood). Langevin dynamics create a stochastic process that eventually converges to samples from the target posterior distribution. It isn't strictly a descent method, it shares some similarities. The negative log-posterior function acts lie a loss function, and the noise term introduces randomness to explore the space of possible volatilities.

References

[1] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.