random forest
bagging

重抽样方法
Bootstrap

统计机器学习方法

	X
1	2
2	3
3	1.1

1, 1, 1 → $\bar{X}_1$

1, 3, 2 → $\bar{X}_2$

⋮

共1000

1, 3, 3 → $\bar{X}_{1000}$

$$\hat{\mu} = \frac{\sum_{i=1}^{1000} \bar{X}_i}{n}.$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{1000} (\hat{\mu} - \bar{X}_i)^2}{999}.$$

Bootstrap

(X, Y)

1 (x_1, y_1)

2 (x_2, y_2)

$\vdots$

100 (x_{100}, y_{100})

1, 2, $\vdots$, 100 $\quad \hat{\alpha}_1$

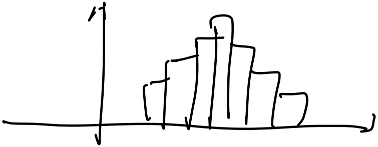
1, 1, 9, $\vdots$, 100 $\quad \hat{\alpha}_2$

$\vdots$

100, 100, $\vdots$, 10 $\quad \hat{\alpha}_{1000}$

$$\hat{\mu}_\alpha = \frac{1}{1000} \sum_{i=1}^{1000} \hat{\alpha}_i$$

$$\hat{\sigma}^2 = \frac{1}{999} \sum_{i=1}^{1000} \frac{(\hat{\alpha}_i - \hat{\mu}_\alpha)^2}{1000}$$



A simple example

$\bar{X}, S_x^2$

$\bar{Y}, S_y^2$

$X = [\ ], \quad Y = [\ ]$

- Suppose that we wish to invest a fixed sum of mone two financial assets that yield returns of X and Y, respectively, where X and Y are random quantities.
- We will invest a fraction α of our money in X, and invest the remaining $1 - \alpha$ in Y.

  $\alpha X + (1-\alpha) Y$

- We wish to choose α to minimize the total risk, or variance, of our investment. In other words, we wan minimize $\mathrm{Var}(\alpha X + (1 - \alpha)Y)$.

$\frac{Q(\alpha)}{d\alpha} = 0$

$Q(\alpha) = \alpha^2 \mathrm{Var}(X) + (1-\alpha)^2 \mathrm{Var}(Y) + 2\alpha(1-\alpha)\,\mathrm{Cov}(X,Y)$

A simple example

- Suppose that we wish to invest a fixed sum of money in two financial assets that yield returns of X and Y, respectively, where X and Y are random quantities.
- We will invest a fraction α of our money in X, and will invest the remaining $1 - \alpha$ in Y.
- We wish to choose α to minimize the total risk, or variance, of our investment. In other words, we want to minimize $\mathrm{Var}(\alpha X + (1 - \alpha)Y)$.

  min (Var)
- One can show that the value that minimizes the risk is given by

$$\boxed{\alpha = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}},$$

$$\hat{\alpha} = \frac{S_Y^2 - S_{XY}}{S_X^2 + S_Y^2 - 2S_{XY}}$$

where $\sigma_X^2 = \mathrm{Var}(X), \sigma_Y^2 = \mathrm{Var}(Y)$, and $\sigma_{XY} = \mathrm{Cov}(X, Y)$.

- But the values of σ_X^2, σ_Y^2, and σ_{XY} are unknown.
- We can compute estimates for these quantities, $\hat{\sigma}_X^2$ ... and $\hat{\sigma}_{XY}$, using a data set that contains measurem... X and Y.
- We can then estimate the value of α that minimiz... variance of our investment using

$$\hat{\alpha} = \frac{\hat{\sigma}_Y^2 - \hat{\sigma}_{XY}}{\hat{\sigma}_X^2 + \hat{\sigma}_Y^2 - 2\hat{\sigma}_{XY}}.$$