notes3 basic of machine learning

1. 数据集
2. 误差分析
3. 代表的机器学习方法
    1. 有监督、线性回归、SVM、决策树、RF
    2. 无监督、聚类、降维（PCA）

machine learning

concept

category

- SupervisedLearning (Category, Regression)
- UnsupervisedLearning (Converge, decrease dimension)
- Reinforcement Learning

dataset

like this, $D = x_1, x_2, \cdots, x_n$ includes *n* samples， x_i is a vector，which shows i sample in the dataset. d is the space dimension

category of dataset

- Trainingset
- Validation set
- Testset

  classic dataset
- 图像分类
    - MNIST http://yann.lecun.com/exdb/mnist/
    - CIFAR-10, CIFAR-100, ImageNet
        - https://www.cs.toronto.edu/~kriz/cifar.html
        - http://www.image-net.org/
    - Large Movie Review Dataset v1.0
        - http://ai.stanford.edu/~amaas/data/sentiment/

- 数据集:https://github.com/researchmm/img2poem

error analysis

over fitting等

lack fitting

general error analysis

$$
\begin{aligned}
\operatorname{Err}(\hat{f}) &= \mathrm{E}\left[(Y-\hat{f}(\mathrm{X}))^2\right] \\
\operatorname{Err}(\hat{f}) &= \mathrm{E}\left[(f(X)+\varepsilon-\hat{f}(\mathrm{X}))^2\right] \\
\operatorname{Err}(\hat{f}) &= \mathrm{E}\left[(f(X)-\hat{f}(\mathrm{X}))^2+2\varepsilon(f(X)-\hat{f}(\mathrm{X}))+\varepsilon^2\right] \\
\operatorname{Err}(\hat{f}) &= \mathrm{E}\left[(E(\hat{f}(\mathrm{X}))-f(X)+\hat{f}(\mathrm{X})-E(\hat{f}(\mathrm{X})))^2\right]+\sigma_\varepsilon^2 \\
\operatorname{Err}(\hat{f}) &= \mathrm{E}[(E(\hat{f}(\mathrm{X}))-f(X))]^2+\mathrm{E}\left[(\hat{f}(\mathrm{X})-E(\hat{f}(\mathrm{X})))^2\right]+\sigma_\varepsilon^2 \\
\operatorname{Err}(\hat{f}) &= \operatorname{Bias}^2(\hat{f})+\operatorname{Var}(\hat{f})+\sigma_\varepsilon^2
\end{aligned}
$$

bias and variance

cross verification

supervised

- 数据集有标记(答案)
- 数据集通常扩展为(x_i,y_i)，其中$y_i \in Y$是 x_i 的标记，Y 是所有标记的集合，称为“标记空间”或“输出空间”
- 监督学习的任务是训练出一个模型用于预测 y 的取值，根据 $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$，训练出函数 f，使得$f(x) \cong y$
- 若预测的值是离散值，如年龄，此类学习任务称为“分类”
- 若预测的值是连续值，如房价，此类学习任务称为“回归”

linear regression

$$
f(x^k)=w_1x_1^k+w_2x_2^k+\cdots+w_mx_m^k+b=\sum_{i=1}^{m}w_ix_i^k+b
$$

$$(w^*, b^*) = argmin_{(w,b)} \sum_{k=1}^{n} (f(x^k) - y^k)^2 = argmin_{(w,b)} \sum_{k=1}^{n} (w^T x^k + b - y^k)^2$$

logistic regression

$$g(f(x^k)) = \begin{cases} 1, \frac{1}{1+e^{-(w^T x^k + b)}} \geq 0.5 \\ 0, otherwise \end{cases}$$

support SVM

$$x = x_0 + \gamma \frac{w}{\|w\|} \tag{1}$$

$$\gamma = \frac{w^T x + b}{\|w\|} = \frac{f(x)}{w} \tag{2}$$

$$\arg\max_{w,b} \arg\min_{x_i \in D} \frac{|w^T x_i + b|}{\sqrt{\sum_{i=1}^{d} w_i^2}}$$
$$s.t. \forall x_i \in D, y_i(w^T x_i + b) \geq 0$$

$\forall x_i \in D, |w^T x_i + b| \geq 1.$

$$\arg\min_{w,b} \frac{1}{2} \sum_{i=1}^{d} w_i^2$$
$$s.t. \forall x_i \in D, |w^T x_i + b| \geq 1$$

decision making tree

Based on data structure Tree

unsupervised

converage

classic algorithm: K-means

lower dimension