北京大学言语听觉研究中心
Speech and Hearing Research Center, Peking University

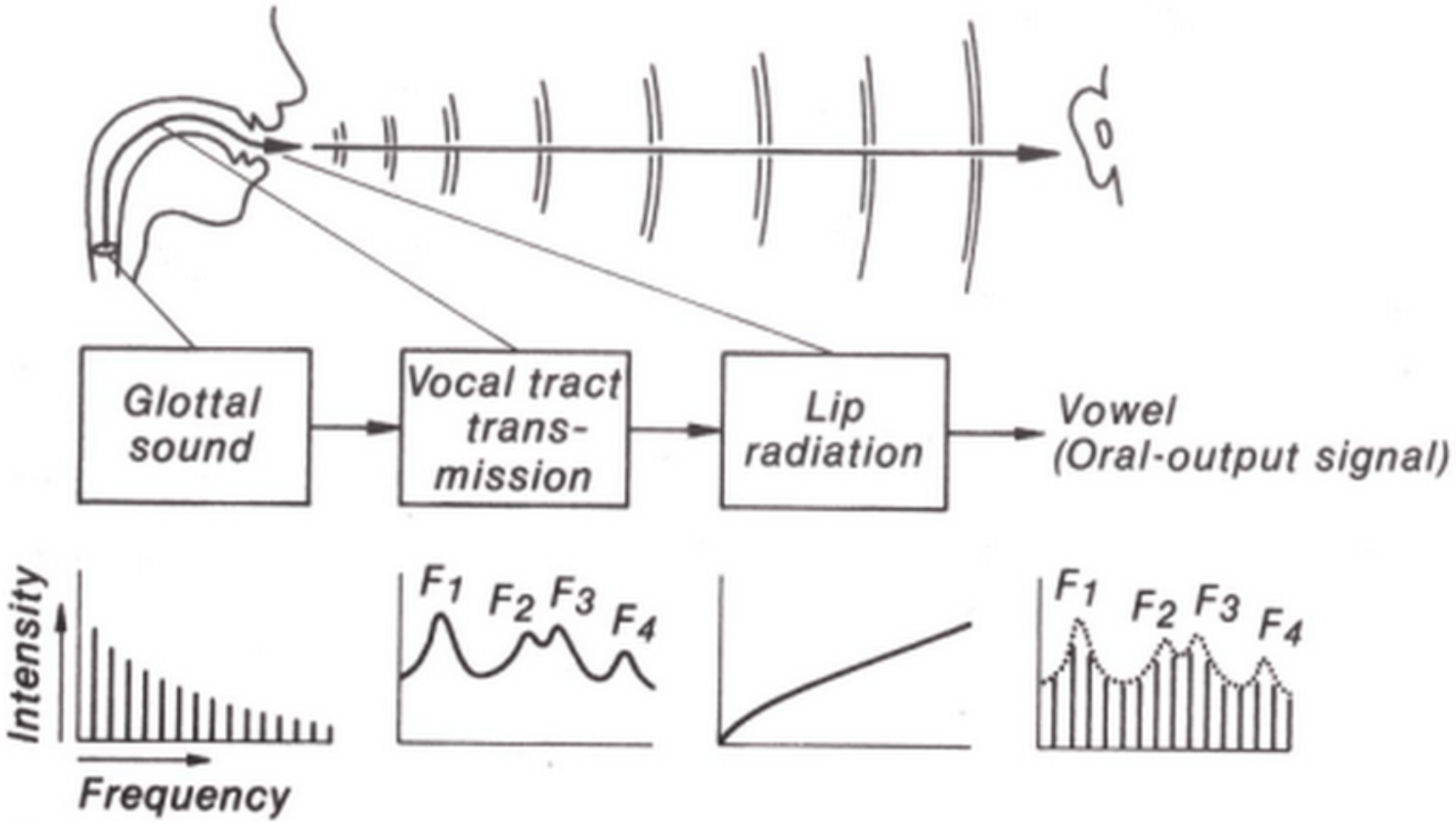
实践一——语音处理综合

陈婧

2021.04.15

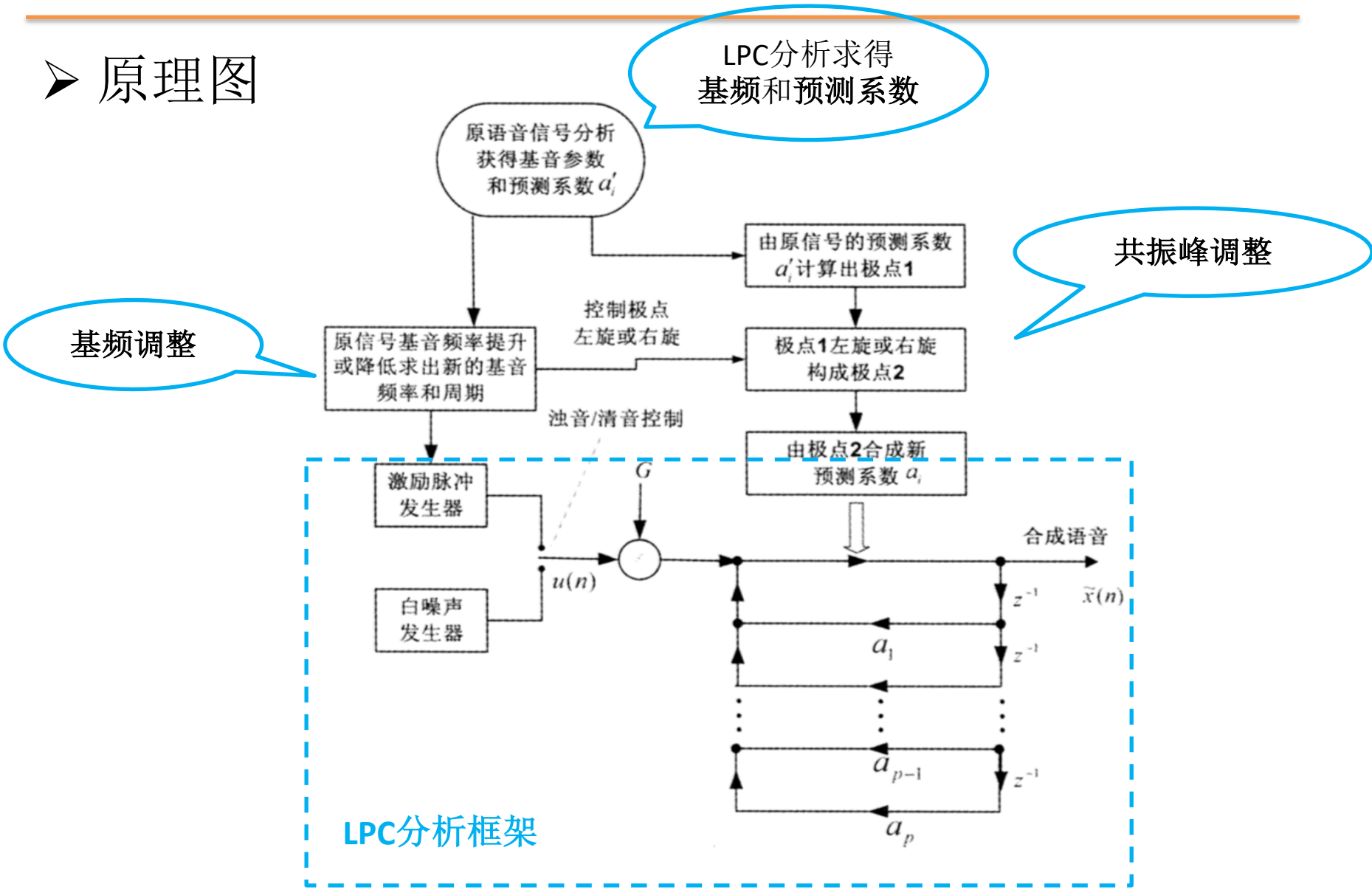
实践1——语音变调

源-滤波器模型

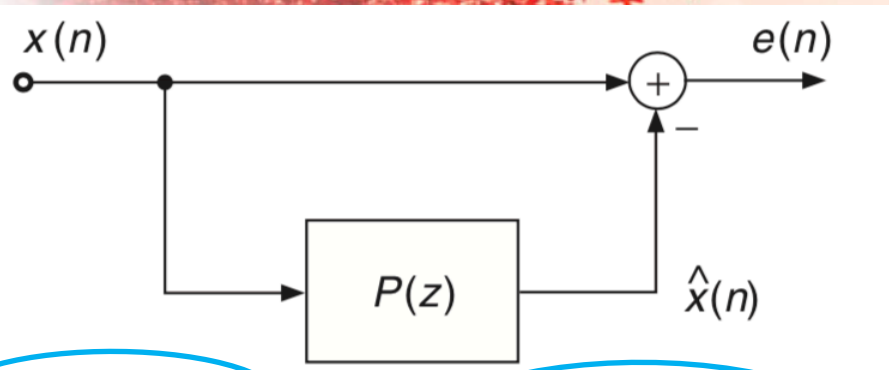


实践1——语音变调

➢ 原理图


LPC分析求得
基频和预测系数
原语音信号分析
获得基音参数
和预测系数 a_i'
由原信号的预测系数
a_i'计算出极点1
共振峰调整
基频调整
控制极点
左旋或右旋
原信号基音频率提升
或降低求出新的基音
频率和周期
极点1左旋或右旋
构成极点2
浊音/清音控制
由极点2合成新
预测系数 a_i
激励脉冲
发生器
G
合成语音
$u(n)$
白噪声
发生器
z^{-1}
$\tilde{x}(n)$
a_1
a_{p-1}
a_p
LPC分析框架

实践1——语音变调



➢ 原理

– LPC分析和基频的关系

激励源

$$x(n) = \sum_{k=1}^{p} a_k x(n-k) + Gu(n)$$

p阶线性预测器的传递函数

$$P(z) = \sum_{k=1}^{p} a_k z^{-k}$$

预测误差

$$\hat{x}(n) = \sum_{k=1}^{p} a_k x(n-k)$$

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{k=1}^{p} a_k x(n-k)$$

$$E(z) = X(z) - \hat{X}(z) = X(z)[1 - P(z)]$$

预测误差滤波器

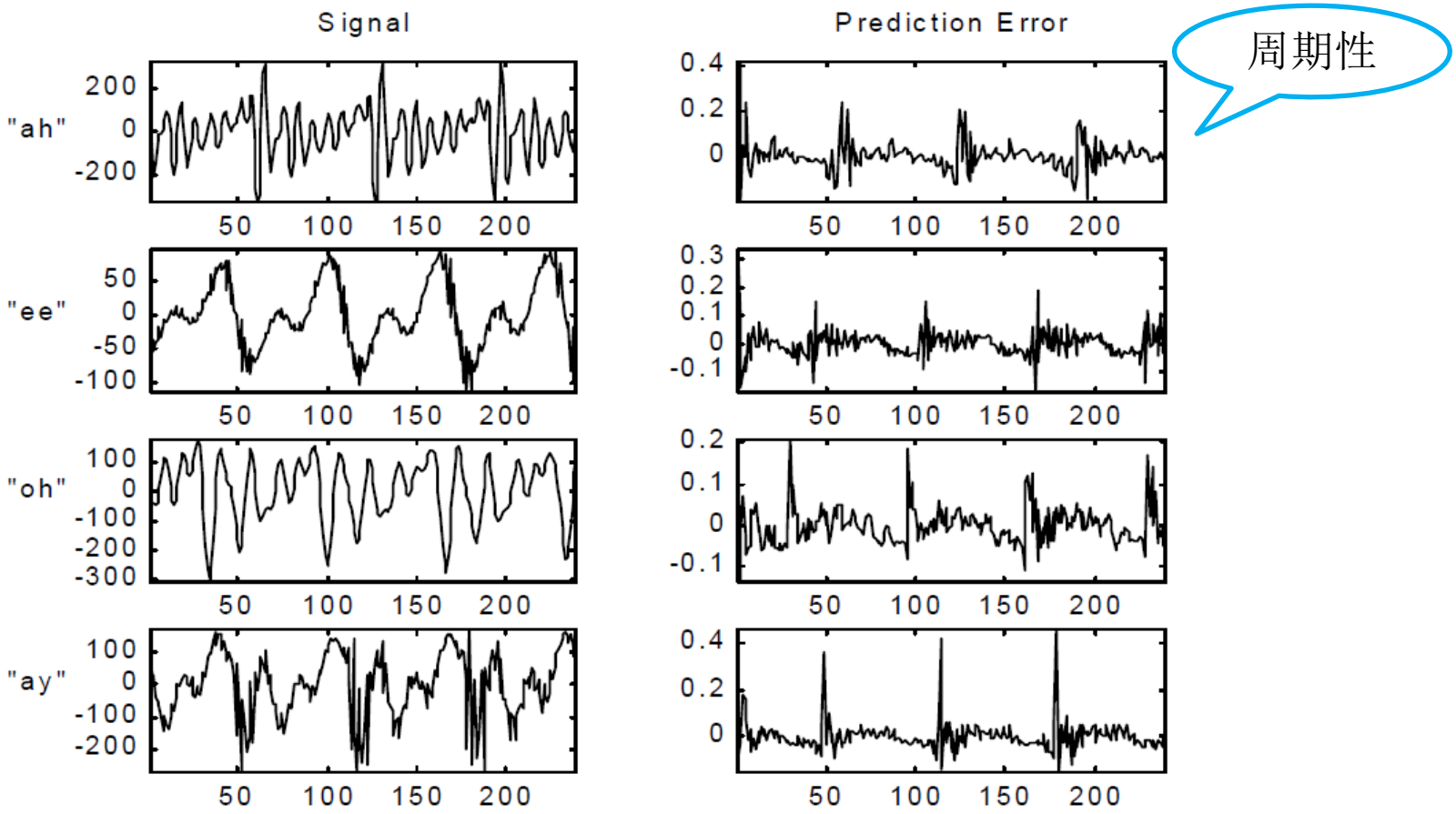
$$A(z) = 1 - P(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}$$

$$E(z) = X(z)A(z)$$

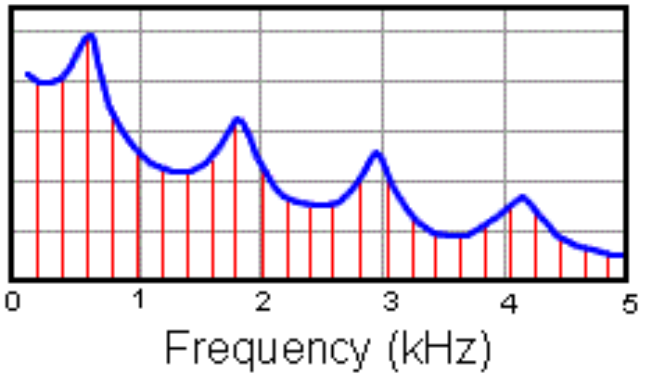
预测误差与激励源对应
激励源中包含基频信息 → 从预测误差中获取基频

实践1——语音变调

- 原理
  - 可从预测误差中获取基频信息



实践1——语音变调



- 原理
  - LPC分析与共振峰的关系

频谱包络

$$E(z) = X(z)A(z)$$

频谱包络中体现了共振峰

$$A(z) = 1 - P(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}$$

方法：通过求预测误差滤波器 $A(z)$ 的根，可以实现对共振峰的估计

设为任意复根值，设与对应的共振峰频率为F_i，带宽为B_i，则

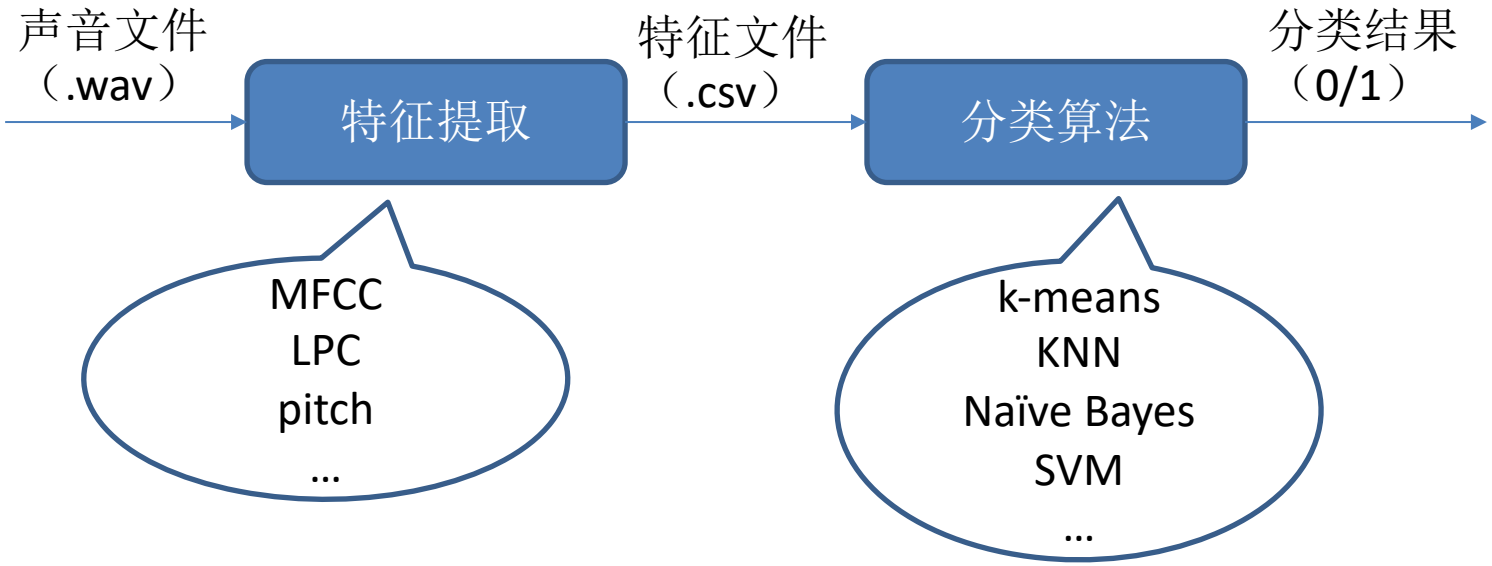
$$2\pi TF_i = \theta_i \qquad F_i = \theta_i/(2\pi T)$$

$$e^{-B_i\pi T} = r_i \qquad B_i = -\ln r_i/\pi T$$

T是采样周期

实践2——声音分类

- 任务描述
  - 给定带有性别标注（female/male）的语音文件，使用机器学习中与监督学习相关的算法，对语音文件进行男声/女声的分类（二分类问题）
- 框架



实践2——声音分类

- 流程
  - 数据获取
    - Kaggle数据集: https://www.kaggle.com/primaryobjects/voicegender#voice.csv
    - 3168条数据，男女各1584条
    - 基于对男女语音段进行合理的声音预处理而得到的语音特征(并不包含原始语音段）
    - 每条数据可视作1×21的向量，前20维为特征值，最后一维为性别标记（以字符串,即male和female进行标注）
    - 可使用matlab读取.csv数据（3168*21的矩阵）
  - 数据处理
    - 量化、缺失值、归一化、训练/测试数据集切分、特征处理、异常值检测……

The Dataset

The following acoustic properties of each voice are measured and included within the CSV:

- **meanfreq**: mean frequency (in kHz)
- **sd**: standard deviation of frequency
- **median**: median frequency (in kHz)
- **Q25**: first quantile (in kHz)
- **Q75**: third quantile (in kHz)
- **IQR**: interquantile range (in kHz)
- **skew**: skewness (see note in specprop description)
- **kurt**: kurtosis (see note in specprop description)
- **sp.ent**: spectral entropy
- **sfm**: spectral flatness
- **mode**: mode frequency
- **centroid**: frequency centroid (see specprop)
- **peakf**: peak frequency (frequency with highest energy)
- **meanfun**: average of fundamental frequency measured across acoustic signal
- **minfun**: minimum fundamental frequency measured across acoustic signal
- **maxfun**: maximum fundamental frequency measured across acoustic signal
- **meandom**: average of dominant frequency measured across acoustic signal
- **mindom**: minimum of dominant frequency measured across acoustic signal
- **maxdom**: maximum of dominant frequency measured across acoustic signal
- **dfrange**: range of dominant frequency measured across acoustic signal
- **modindx**: modulation index. Calculated as the accumulated absolute difference between adjacent measurements of fundamental frequencies divided by the frequency range
- **label**: male or female

实践2——声音分类

- 流程（续）
  - 建模

Logistic Regression

97% / 98%

CART

96% / 97%

Random Forest

100% / 98%

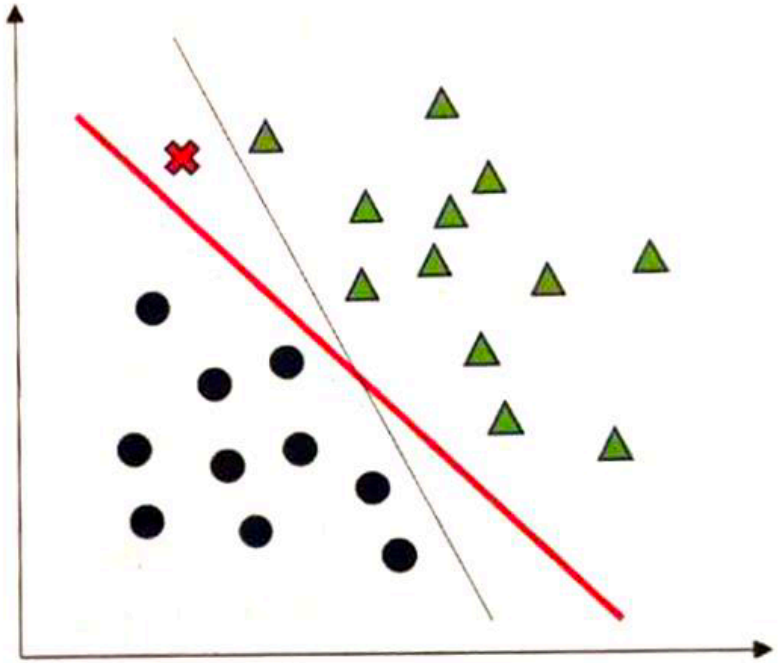
SVM

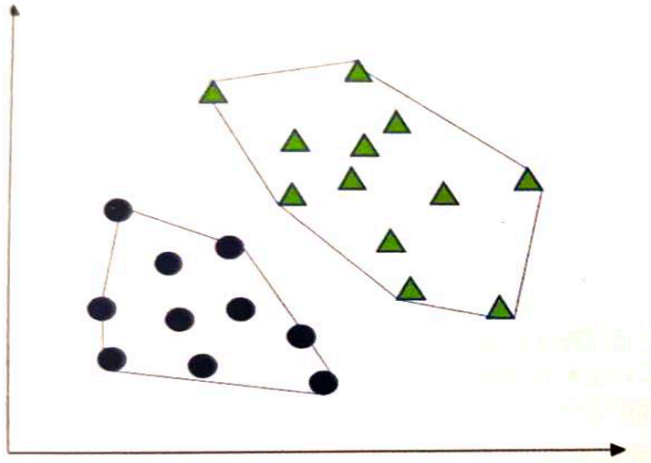
100% / 99%

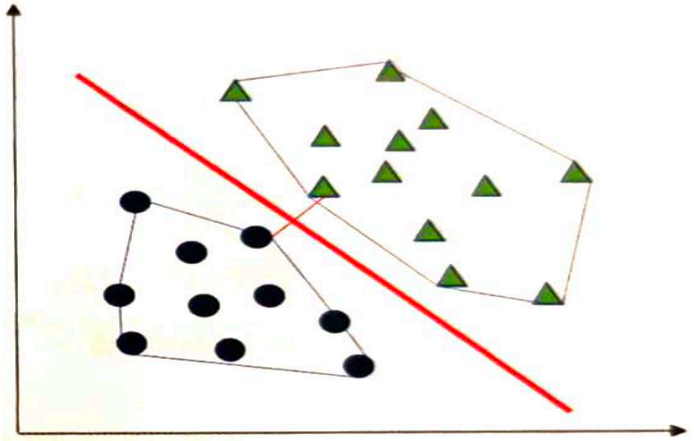
XGBoost

100% / 99%

SVM Support Vector Machine



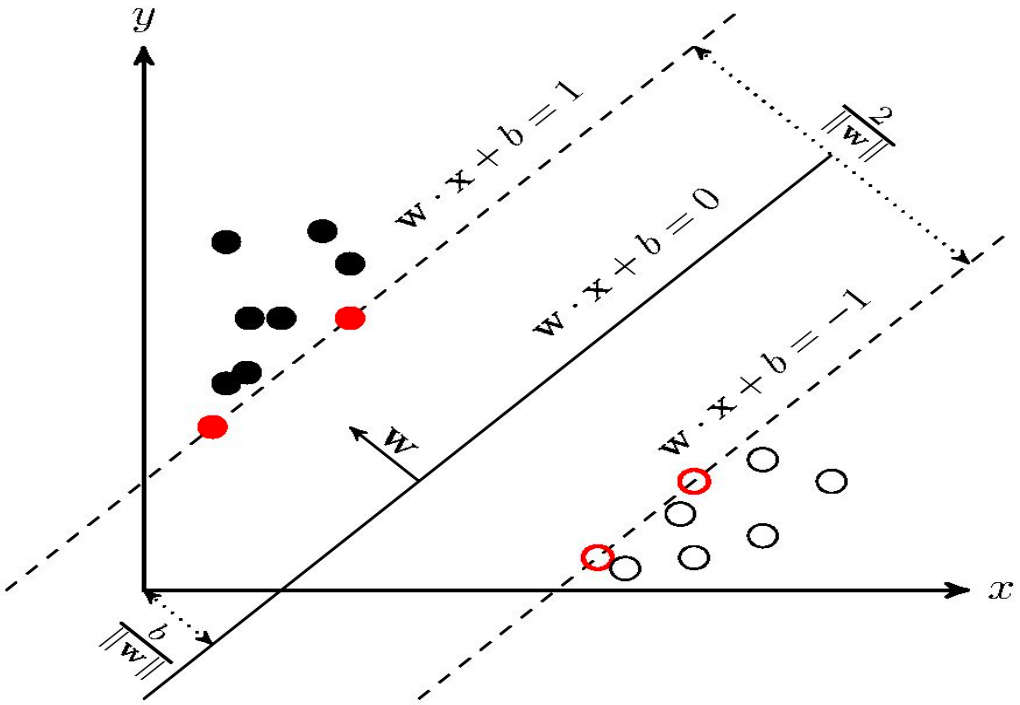




SVM Support Vector Machine

SVM的基本思想

- 求解能够正确划分训练数据集并且几何间隔最大的分离超平面
- 图中$\boldsymbol{w}\cdot x+\ b\ =\ 0$即为分离超平面



实践2——声音分类

- 流程（续）
  - 验证和测试
    - 如何划分数据集
    - 评价准则：准确率、召回率 …
  - 模型改进
    - 各种参数的调整对结果的影响
  - 结果分析
    - 列表、可视化 …

实验报告要求

- 报告大纲
  - 基本原理
  - 方法
  - 实验结果与讨论
  - 总结

- 提交时限
  - 4.15-4.29