2020270026 应用统计硕士 王姿文

1. Boosting: from Weak to Strong

Problem 1

此題是想介紹$Boosting\ Simple\ Thresholding-Based\ Decision\ Stumps$，因為$Decision\ Stumps$是一種弱分類器，它的深度只有一層，意思是僅用一個feature去分類，但這樣分類效果肯定不好，因此本題結合$Boosting+Thresholding-Based$的概念，來使原本的目標函數$argmin\frac{1}{m}\sum_m \mathbb{I}\{y_i \neq \phi_n x_i\}$改為

$$\begin{cases} \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} \leq \frac{1}{2}-\gamma, \\ \sum_m P_i \mathbb{I}\{y^{(i)} \neq \phi_{s,-} x^{(i)}\} \leq \frac{1}{2}-\gamma \end{cases}$$

$$\forall \{(x^{(i),y^{(i)}})\}_{i=1}^m \sim \{p^{(i)}\}_{i=1}^m,\ y \in \{-1,1\},\ \gamma > 0,$$
$$\mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} = \begin{cases} 1, y^{(i)} \neq \phi_{s,+} x^{(i)} \\ 0, otherwise \end{cases}$$

$$\phi_{s,+} x^{(i)} = \begin{cases} 1, x \geq s \\ -1, x < s \end{cases}$$

$$\phi_{s,+} x^{(i)} = -\phi_{s,-} x^{(i)},\ J_t \leq \sqrt{1-4\gamma^2} J_{t-1}$$
$$x^{(1)} > x^{(2)} > \ldots > x^{(m)}$$

且可使得$J_t < \frac{1}{m}$，亦即$zero\ error\ rate$

1

欲证

$$\begin{cases} \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} = \frac{1}{2} - \frac{1}{2}(\sum_{i=1}^{m_0(s)} y^{(i)} p_i - \sum_{i=m_0(s)+1}^{m} y^{(i)} p_i), \\ \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,-} x^{(i)}\} = \frac{1}{2} - \frac{1}{2}(\sum_{i=m_0(s)+1}^{m} y^{(i)} p_i - \sum_{i=1}^{m_0(s)} y^{(i)} p_i) \end{cases}$$

$$\forall s,\ m_0(s) \in \{0, 1, .., m\}$$

下证：

$\because \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} = \begin{cases} 1, y^{(i)} \neq \phi_{s,+} x^{(i)} \\ 0, otherwise \end{cases}$

$\therefore \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} = 1\ when\ \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} = \mathbb{I}\{\phi_{s,+} x^{(i)} = 1, y^{(i)} = -1\} + \mathbb{I}\{\phi_{s,+} x^{(i)} = -1, y^{(i)} = 1$

$= \mathbb{I}\{x^{(i)} \geq s, y^{(i)} = -1\} + \mathbb{I}\{x^{(i)} < s, y^{(i)} = 1\}$

$\Rightarrow \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} = \sum_m p_i \mathbb{I}\{x^{(i)} \geq s, y^{(i)} = -1\} + \sum_m p_i \mathbb{I}\{x^{(i)} < s, y^{(i)} = 1\}$

又 $\because \mathbb{I}\{y = -1\} = \frac{1}{2} - \frac{y}{2},\ \mathbb{I}\{y = -1\} = \frac{1}{2} + \frac{y}{2}$

$\therefore \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\}$

$= \sum_m p_i \mathbb{I}\{x^{(i)} \geq s, y^{(i)} = -1\} + \sum_m p_i \mathbb{I}\{x^{(i)} < s, y^{(i)} = 1\}$

$= \sum_{i=1}^{m_0(s)} p_i \mathbb{I}\{y = -1\} + \sum_{i=m_0(s)+1}^{m} p_i \mathbb{I}\{y = 1\}$

$= \frac{1}{2} - \frac{1}{2} \sum_{i=1}^{m_0(s)} y^{(i)} p_i + \frac{1}{2} \sum_{i=m_0(s)+1}^{m} y^{(i)} p_i$

$= \frac{1}{2} - \frac{1}{2}(\sum_{i=1}^{m_0(s)} y^{(i)} p_i - \sum_{i=m_0(s)+1}^{m} y^{(i)} p_i)$

上式得证

接着用同样概念可得

$\sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,-} x^{(i)}\} = \sum_m p_i \mathbb{I}\{x^{(i)} \geq s, y^{(i)} = 1\} + \sum_m p_i \mathbb{I}\{x^{(i)} < s, y^{(i)} = -1\}$

$= \sum_{i=1}^{m_0(s)} p_i \mathbb{I}\{y = 1\} + \sum_{i=m_0(s)+1}^{m} p_i \mathbb{I}\{y = -1\}$

$= \frac{1}{2} + \frac{1}{2} \sum_{i=1}^{m_0(s)} y^{(i)} p_i - \frac{1}{2} \sum_{i=m_0(s)+1}^{m} y^{(i)} p_i$

$= \frac{1}{2} - \frac{1}{2}(\sum_{i=m_0(s)+1}^{m} y^{(i)} p_i - \sum_{i=1}^{m_0(s)} y^{(i)} p_i)$

下式得证

2

欲证：

$$max_{m_0} \mid f(m_0) \mid \geq 2\gamma, \forall \gamma = \frac{1}{2m},\ f(m_0) = \sum_{i=1}^{m_0} y^{(i)} p_i - \sum_{m_0+1}^{m} y^{(i)} p_i,\ m_0 \in \{0, \ldots, m\}$$

下证：

$$\because \begin{cases} \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} \leq \frac{1}{2} - \gamma, \\ \sum_m P_i \mathbb{I}\{y^{(i)} \neq \phi_{s,-} x^{(i)}\} \leq \frac{1}{2} - \gamma \end{cases}$$

$$\therefore \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_s x^{(i)}\} \leq 1 - 2\gamma - (1)$$

$$\text{又} \because \begin{cases} \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} = \frac{1}{2} - \frac{1}{2}(\sum_{i=1}^{m_0(s)} y^{(i)} p_i - \sum_{i=m_0(s)+1}^{m} y^{(i)} p_i), \\ \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,-} x^{(i)}\} = \frac{1}{2} - \frac{1}{2}(\sum_{i=m_0(s)+1}^{m} y^{(i)} p_i - \sum_{i=1}^{m_0(s)} y^{(i)} p_i) \end{cases}$$

$$\therefore \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_s x^{(i)}\} = 1 - \frac{1}{2} p_i \mathbb{I}\{y^{(i)=-1}\} + \frac{1}{2} p_i \mathbb{I}\{y^{(i)=1}\} = 1 - \mid p_i y^{(i)} \mid -(2)$$

$$By\ (1), (2),\ 1 - \mid p_i y^{(i)} \mid \leq 1 - 2\gamma$$

$$\Rightarrow \mid p_i y^{(i)} \mid \geq 2\gamma$$

$$\Rightarrow \mid p_i y^{(i)} \geq \frac{1}{m} - (3)$$

$$By\ (1), (2), (3),$$

$$f(m_0) - f(m_0 + 1) = \sum_{i=1}^{m_0} y^i p_i - \sum_{i=1}^{m} y^i p_i - \sum_{i=1}^{m_0+1} y^i p_i + \sum_{i=m_0+2}^{m} y^i p_i$$

$$= -2y^{m_0+1} p_{m_0+1}$$

$$= -2 * y^{m_0+1} p_{m_0+1}$$

$$\Rightarrow \mid f(m_0) - f(m_0 + 1) \mid = 2 * \mid y^{m_0+1} p_{m_0+1} \mid \geq 2 * \frac{1}{m}$$

$$\Rightarrow \mid f(m_0) - f(m_0 + 1) \mid \geq \frac{2}{m}$$

$$\because 2 * max_{m_0} \mid f(m_0) \mid \geq \mid f(m_0) \mid + \mid f(m_0 + 1) \mid \geq \mid f(m_0) - f(m_0 + 1) \mid \geq \frac{2}{m}$$

$$\therefore 2 * max_{m_0} \mid f(m_0) \mid \geq \mid f(m_0) - f(m_0 + 1) \mid \geq \frac{2}{m}$$

$$\Rightarrow max_{m_0} \mid \geq \frac{1}{m}$$

$$\Rightarrow max_{m_0} \mid f(m_0) \mid \geq 2\gamma, \forall \gamma = \frac{1}{2m}$$

3

已知$\begin{cases} \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} \leq \frac{1}{2} - \gamma, \\ \sum_m P_i \mathbb{I}\{y^{(i)} \neq \phi_{s,-} x^{(i)}\} \leq \frac{1}{2} - \gamma \end{cases} \forall \gamma > 0$

且$\begin{cases} \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} = \frac{1}{2} - \frac{1}{2}(\sum_{i=1}^{m_0(s)} y^{(i)} p_i - \sum_{i=m_0(s)+1}^{m} y^{(i)} p_i), \\ \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,-} x^{(i)}\} = \frac{1}{2} - \frac{1}{2}(\sum_{i=m_0(s)+1}^{m} y^{(i)} p_i - \sum_{i=1}^{m_0(s)} y^{(i)} p_i) \end{cases}$

根据上述知
$\frac{1}{2} - \frac{1}{2} \mid (\sum_{i=m_0(s)+1}^{m} y^{(i)} p_i - \sum_{i=1}^{m_0(s)} y^{(i)} p_i) \mid \leq \frac{1}{2} - \gamma$
$\Rightarrow \mid (\sum_{i=m_0(s)+1}^{m} y^{(i)} p_i - \sum_{i=1}^{m_0(s)} y^{(i)} p_i) \mid \geq 2\gamma$

又知道$max_{m0} \mid f(m_0) \mid \geq \frac{1}{m}, \forall f(m_0) = \sum_{i=1}^{m_0} y^{(i)} p_i - \sum_{m_0+1}^{m} y^{(i)} p_i,\ m_0 \in \{0, \ldots, m\}$

$\Rightarrow \mid (\sum_{i=m_0(s)+1}^{m} y^{(i)} p_i - \sum_{i=1}^{m_0(s)} y^{(i)} p_i) \mid \geq \frac{1}{m}$
$\Rightarrow \gamma = \frac{1}{2m}$

故给定$\begin{cases} \sum_m p_i \mathbb{I}\{y^{(i)} \neq \phi_{s,+} x^{(i)}\} \leq \frac{1}{2} - \gamma, \\ \sum_m P_i \mathbb{I}\{y^{(i)} \neq \phi_{s,-} x^{(i)}\} \leq \frac{1}{2} - \gamma \end{cases}$，可以保证margin$\gamma = \frac{1}{2m}$，并且达到
$J_t < \frac{1}{m}$，亦即$zero\ error\ rate$

接著根據$J_t \leq \sqrt{1 - 4\gamma^2} J_{t-1}$來求upper bound on the number of thresholded decision stumps
$\because \gamma = \frac{1}{2m}$
$\therefore J_t \leq \sqrt{1 - \frac{1}{m^2}} J_{t-1}$
$\Rightarrow J_t \leq \sqrt{1 - \frac{1}{m^2}}^t J_0$
$\Rightarrow J_t \leq \sqrt{1 - \frac{1}{m^2}}^t$
$\Rightarrow J_t \leq \sqrt{\frac{m^2-1}{m^2}}^t$
$\Rightarrow J_t \leq (\frac{\sqrt{m^2-1}}{m})^t$
又$\because J_t \leq \frac{1}{m}$
$\Rightarrow (\frac{\sqrt{m^2-1}}{m})^t \geq \frac{1}{m}$
$\Rightarrow t \ln(\frac{\sqrt{m^2-1}}{m}) \geq -1 \ln m$
$\Rightarrow t \ln(\frac{m}{\sqrt{m^2-1}}) \leq \ln m$
$\Rightarrow t \leq \frac{\ln m}{\ln(\frac{m}{\sqrt{m^2-1}})}$
upper bound on the number of thresholded decision stumps $= \frac{\ln m}{\ln(\frac{m}{\sqrt{m^2-1}})}$

---

2. Deep Neural Networks: Have a Try

在A Neural Networks Playground中，有需多参数跟配置可以调整，由于本题仅提及 `learning rate` 、 `activation function` 、 `n of hidden layer` 、 `regularization` ，因此尝试调参的过程中我便没去调整 `input x` （此网页可以调整成$x_i^2, sin(x_i)$） 。
首先设定：

- Ratoi of training to test data: 50%
- Noise: 10
- Batch size: 10
- Problem type: classification

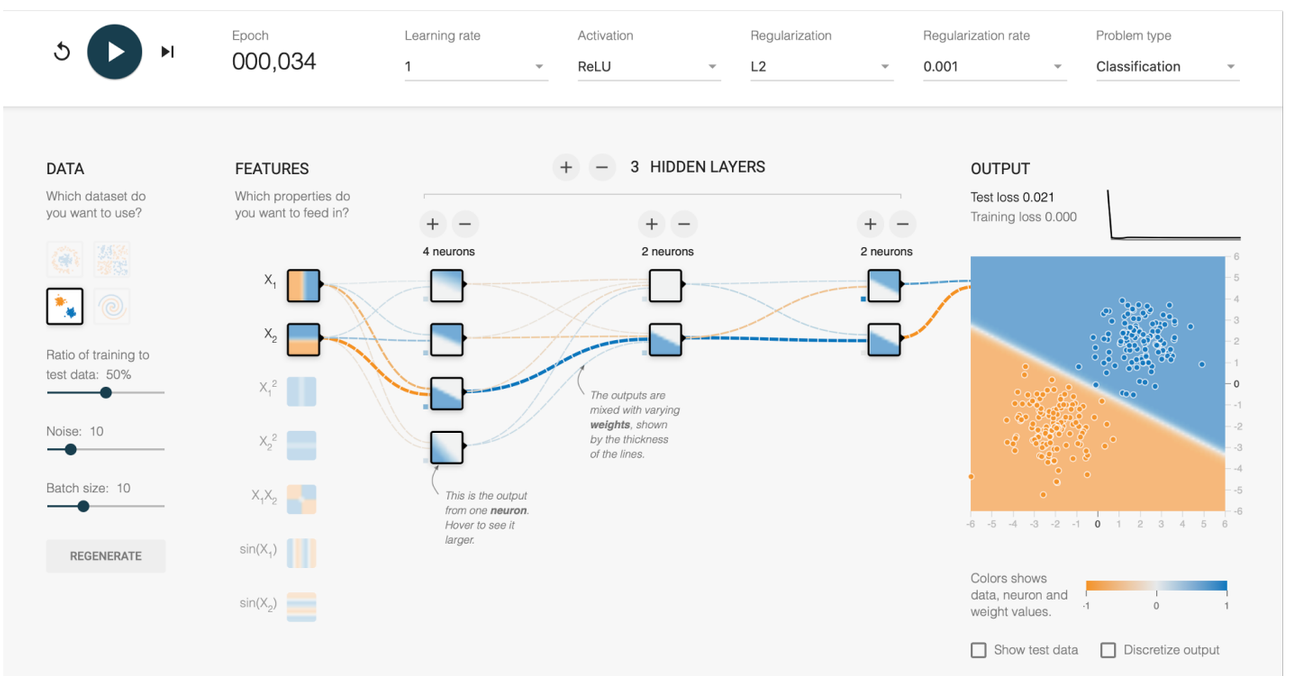
接着依照以下各个参数的影响来调整：

- `learning rate` : 梯度下降的学习率，设定高虽然速度快但由于太高导致步数太大可能导致找不到最优解或无法收敛，反之设定小虽然可以找到最优解但要考量整体模型训练时常来设定
- `activation function` : 模型所使用的function，本题有 `ReLU` 、 `Tanh` 、 `Sigmoid` 、 `Linear` ，应该依照数据的特性来选择
- `n of hidden layer` : 多个隐藏层适合nonlinear-function，虽然越多层照理说会训练得越好，但要考虑训练时长以及overfitting的问题
- `regularization` : 为了防止overfitting而设定，有 `L1` 、 `L2` ，其中 `L1` 会导致某些参数可能变为0，所以有feature selection的效果

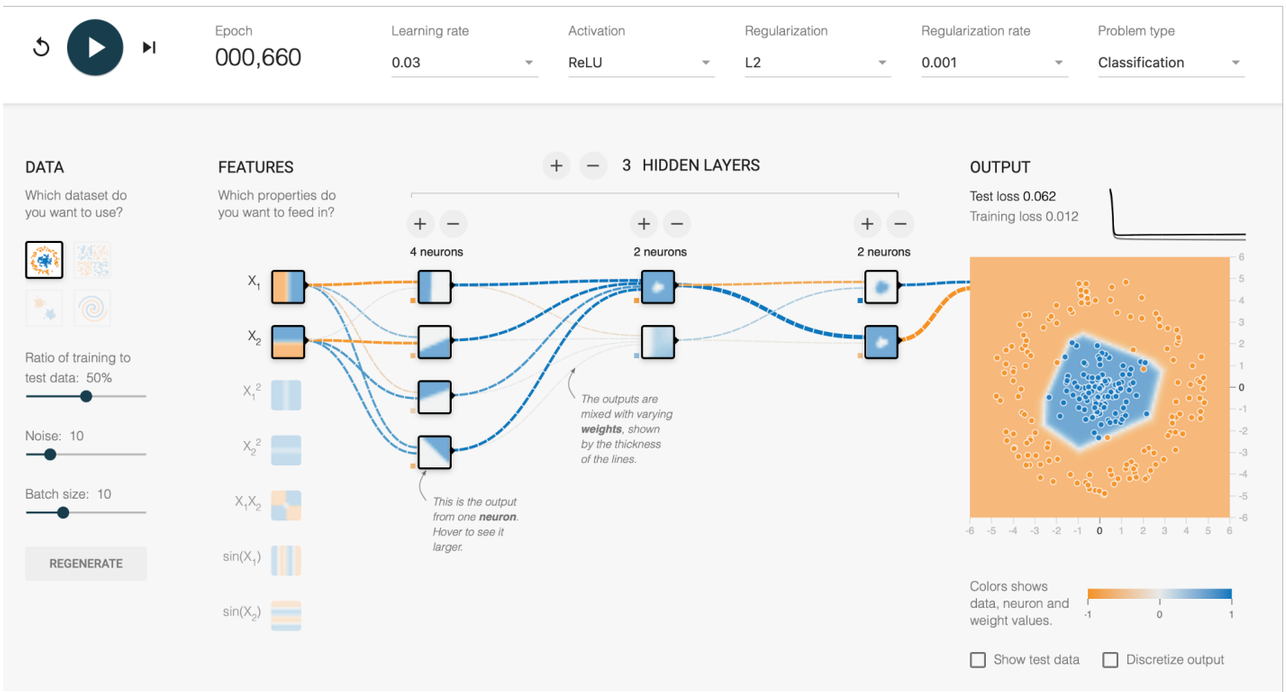
判断参数配置好坏：
可以用 `Epoch` 、 `Converage rate` (= `Train Loss` 、 `Test Loss`)来判断

1 Best Configuration of Gaussian, Circle, XOR and Spiral

- Gaussian
  `learning rate` = 1、 `activation function` = `ReLU` 、 `regulariztion` = `L2` 、 `regulariztion rate` =0.001、 `n of hidden layer` =3

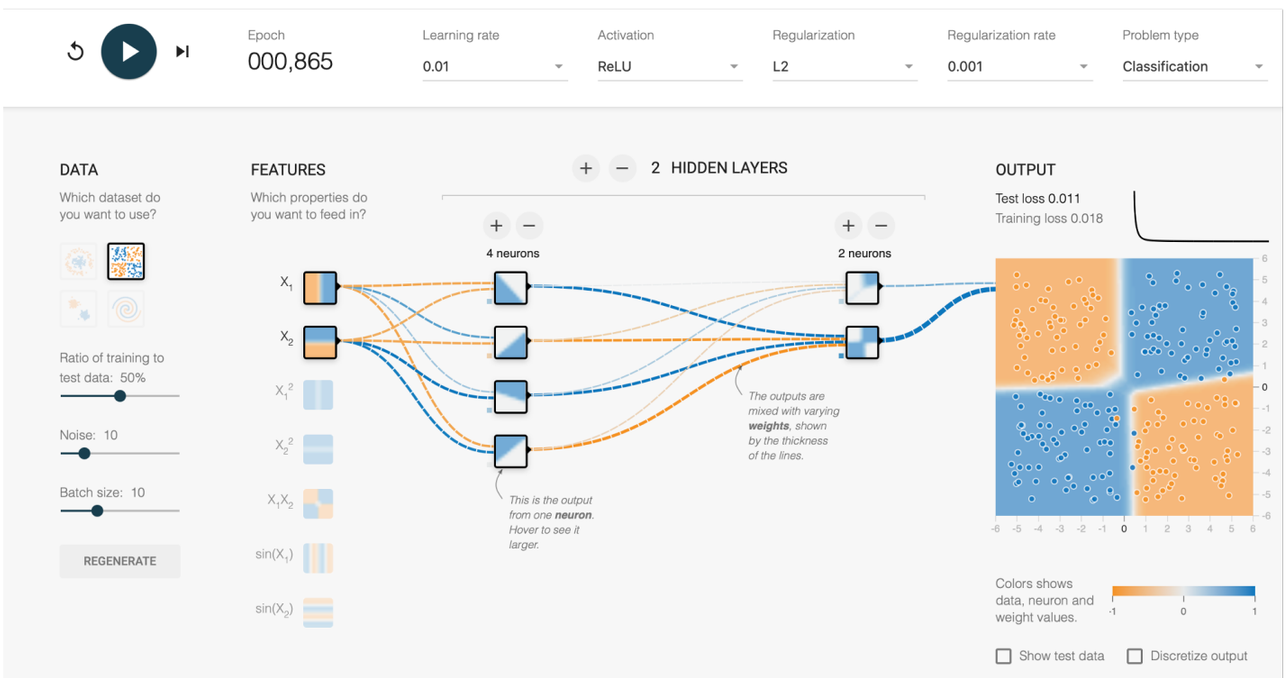


- Circle
  `learning rate` = 0.03、 `activation function` = `ReLU` 、 `regulariztion` = `L2` 、 `regulariztion rate` =0.001、 `n of hidden layer` =3



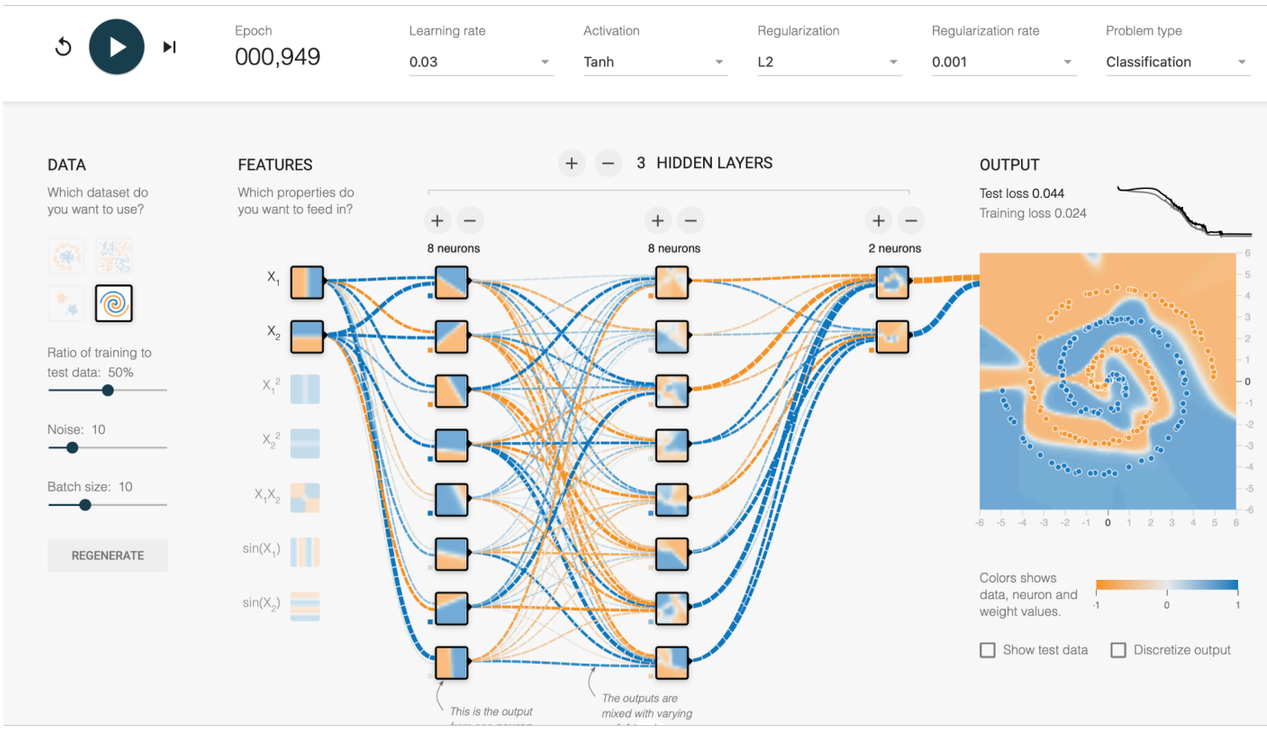
- XOR

  learning rate = 0.01、activation function = ReLU 、regulariztion = L2 、regulariztion rate =0.001、n of hidden layer =2



- Spiral

  learning rate = 0.03、activation function = Tanh 、regulariztion = L2 、regulariztion rate =0.001、n of hidden layer =3 (只有**Spiral**有调整以增加隐藏层内的节点)


Epoch
000,949
Learning rate
0.03
Activation
Tanh
Regularization
L2
Regularization rate
0.001
Problem type
Classification
DATA
Which dataset do you want to use?
Ratio of training to test data: 50%
Noise: 10
Batch size: 10
REGENERATE
FEATURES
Which properties do you want to feed in?
X1
X2
X1²
X2²
X1X2
sin(X1)
sin(X2)
3 HIDDEN LAYERS
8 neurons
8 neurons
2 neurons
This is the output
The outputs are mixed with varying
OUTPUT
Test loss 0.044
Training loss 0.024
Colors shows data, neuron and weight values.
Show test data
Discretize output

2 Findings that how the parameters influence the performance and convergence rate

- 整体来说：
  - learning rate：
    - 上升：可能无法收敛或是 Epoch 下降， Converage rate 可能不变、下降、不稳定
    - 下降： Epoch 上升， Converage rate 可能上升
  - activation function :虽然不同方法适用于不同数据，但也会依照其他参数已想该方法的表现
    - ReLU ：都适用
    - Tanh ：都适用
    - Sigmoid ：只有 Spiral 不适用
    - Linear ：只有在 Gaussian 比较适用，其余因为数据分布难以用线性来当界线区分因此表现不佳
  - n of hidden layer：
    - 上升：训练结果照理说更好， Epoch 上升， Converage rate 下降，对于一些无法收敛的数据，隐藏层数上升可能可以帮助收敛
    - 下降：训练结果照理说没多层一些来得好， Epoch 下降， Converage rate 可能上升
  - regularization：
    - L1 ： Epoch 可能上升、 Train Loss 下降
    - L2 ： Epoch 可能上升、 Train Loss 下降
  - regularization rate：
    - 上升则 Epoch 、 Train Loss 下降， Test Loss 不一定
- 分数据来看
  - Gaussian：
    - activation function 优到劣： ReLU > Tanh = Linear > Sigmoid
    - n of hidden layer ：上升反而(3->4)， Converage rate 下降
  - Circle：
    - learning rate 影响较大
    - LR=1， activation function 优到劣:ReLU=Sigmoid>Linear，其中Tanh表现最佳但 Train Loss 、 Test Loss 波动大（难以收敛）
    - LR=0.03， activation function 优到劣:ReLU>Tanh>Sigmoid>Linear
  - XOR：
    - learning rate 影响较大，且 Tanh ReLU 的 Train Loss 、 Test Loss 波动严重（难以收敛），而在 learning rate 较大时是 Sigmoid 表现最好，但 Sigmoid 相对其他方法的表现会随着 learning rate 的减少而下降
    - LR=1， activation function 优到劣: Sigmoid>Tanh>ReLU>>Linear
    - LR=0.1， activation function 优到劣: Sigmoid稳定=Tanh=ReLU波动严重
    - LR=0.01， activation function 优到劣: ReLU>Tanh>>Sigmoid>>Linear
  - Spiral：
    - 只有 Tanh 和 ReLU 可以使 Converage rate 下降，且还需要增加隐藏层内的节点否则 Train Loss 、 Test Loss 难以下降

3 Principle

这次的尝试中，我会先选定一个类型的data，并按照下述方法来选择最好的参数配置：

1. 固定 learning rate ，选定不同 activation function ，查看 Epoch 和 Converage rate 状况
2. 大致了解哪些 activation function 较好后，调整 learning rate ，查看 Epoch 和 Converage rate 状况
3. 选定 activation function 和 learning rate ，接着固定 regularization rate ，查看 Epoch 和 Converage rate 状况以决定要用 L1 或 L2
4. 选定 activation function 、 learning rate 、 regularization ，最后调整 regularization rate ，查看 Epoch 和 Converage rate 状况
5. 产生最优参数配置

3. Clustering: Mixture of Multinomials

Problem 3

MLE for multinomial

已知$\mu = (\mu)_{i=1}^{d}$,
$P(\mathbf{x}|\mu) = \frac{n!}{\prod_i x_i!} \prod_i \mu_i^{x_i},\ \forall\ i = 1, \ldots, d\ x_i \in \mathbb{N},\ \sum_i x_i = n, 0 < \mu_i < 1,\ \sum_i \mu_i = 1$

下用MLE求$\hat{\mu}$:

$L(\mu) = \prod_i P(x_i|\mu) = P(\mathbf{x}|\mu) = \frac{n!}{\prod_i x_i!} \prod_i \mu_i^{x_i}$
$\Rightarrow \ln L(\mu) = \ln n! - \sum_i \ln x_i! + \sum_i x_i \ln \mu_i$

$\because \sum_i \mu_i = 1\ \therefore$ *By Lagrange multiplier*
$\ln' L(\mu, \lambda) = \ln L(\mu) + \lambda(1 - \sum_i \mu_i)$

欲求$argmax_p L(\mu, \lambda)$

$\frac{d}{d\mu_i} \ln' L(\mu, \lambda) = \frac{d}{d\mu_i} \ln L(\mu) + \frac{d}{d\mu_i} \lambda(1 - \sum_i \mu_i) = 0$
$\Rightarrow \frac{d}{d\mu_i}(\ln n! - \sum_i \ln x_i! + \sum_i x_i \ln \mu_i) + \frac{d}{d\mu_i} \lambda(1 - \sum_i \mu_i) = 0$
$\Rightarrow \frac{x_i}{\mu_i} - \lambda = 0$
$\Rightarrow x_i = \lambda \mu_i$
$\Rightarrow \hat{\mu_i} = \frac{x_i}{\lambda}$

又$\because \sum_i \mu_i = \sum_i \frac{x_i}{\lambda}$
$\Rightarrow 1 = \sum_i \frac{x_i}{\lambda}$
$\Rightarrow 1 = \frac{1}{\lambda} \sum_i x_i$
$\Rightarrow \lambda = \sum_i x_i = n$
$\Rightarrow \lambda = n$
$\Rightarrow \hat{\mu_i} = \frac{x_i}{n}$

Problem 4

Derive EM for mixture of multinomials

EM算法是用已知的变量来迭代估计出未知的变量：

1. 初始化分布参数
2. 重复EM步骤直到收敛
   - E: 求出未知参数的期望估计，以给定参数缺失的数据
   - M: 将E代入MLE以优化参数

以下为EM步骤的实现方式： 给定数据集，假设样本间相互独立，我们想要拟合模型$p(x;\theta)$ 到数据的参数。根据分布我们可以得到如下似然函数：

$L(\theta) = \sum_i \ln p(x_i;\theta) = \sum_i \ln \sum_z p(x_i, z;\theta),\ \forall z_i \sim q_i(z),\ \sum_z q_i(z) = 1,\ q_i(z) \geq 0,\ z$是隐变量

$By\ Jenson\ Inequality\ ,\ E[f(X)] \geq f[E(X)],\ \forall f(x) = \ln \frac{p(x;\theta)}{q_i(z)}$是凹函数

$\Rightarrow L(\theta) = \sum_i \ln \sum_z p(x_i, z;\theta) = \sum_i \ln \sum_z q_i(z) \frac{p(x_i,z;\theta)}{q_i(z)} \geq \sum_i \sum_z q_i(z) \ln \frac{p(x_i,z;\theta)}{q_i(z)} - (1)$

接着推导 $q_i(z)$，$\because$ 在$X = E[X]$时，X为常数等式才成立，$\therefore \frac{p(x_i,z;\theta)}{q_i(z)} = c$

$\Rightarrow \forall \sum_z q_i(z) = 1,\ \sum_z p(x_i, z, \theta) = c$

$\Rightarrow q_i(z) = \frac{p(x_i,z,\theta)}{\sum_z p(x_i,z,\theta)} = \frac{p(x_i,z,\theta)}{p(x_i,\theta)=p(z|x_i,\theta)} - (2)$

这就是EM算法的实现方式(1)=E,(2)=M，接着下将mixture of multinomials应用于EM

已知T，且$P(C_d = k) = \pi_k,\ k = 1, 2, .., k$

$\mu_{\mathbb{k}} = (\mu_{1k} \ldots \mu wk)$

$p(d|C_d = k) = \frac{n_d!}{\prod_w T_{dw}!} \prod_w \mu_{wk}^{T_{dw}}, \forall\ n_d = \sum_w T_{dw}$

$\Rightarrow p(d) = \sum_{k=1}^{K} p(d|C_d = k)p(C_d = k) = \frac{n_d!}{\prod_w T_{dw}!} \sum_k \pi_k \prod_w \mu_{wk}^{T_{dw}}$

(1) $L(\mu, \pi) = \sum_d \ln p(d; C_d) = \sum_d \ln \sum_k p(d) \geq \sum_d \sum_k q_{dk} \ln \frac{p(d)}{q_{dk}}$

$= \sum_d \sum_k q_{dk} \ln \frac{p(d|C_d=k)p(C_d=k)}{q_{dk}}$

$= \sum_d \sum_k [q_{dk} \ln p(d|C_d = k) + q_{dk} \ln \pi_k - q_{dk} \ln q_{dk}]$

$= \sum_d \sum_k q_{dk} [\ln n_d! - \sum_d \ln T_{dw}! + \sum_d T_{dw} \ln \mu_{wk} + \ln \pi_k - \ln q_{dk}]$

$= B(\mu, \pi)$

(2)$\because \sum_{w,k} \mu_{wk} = 1$

$\Rightarrow By\ Lagrange\ multiplier,\ B^{'}(\mu) = B(\mu) + \lambda(1 - \sum_{w,k} \mu_{wk})$

$\Rightarrow \frac{dB^{'}(\mu)}{d\mu_{wk}} = \frac{dB(\mu)}{d\mu_{wk}} - \lambda = 0$

$\Rightarrow \frac{d}{d\mu_{wk}} \sum_d \sum_k q_{dk} \sum_d T_{dw} \ln \mu_{wk} = \lambda$

$\Rightarrow \hat{\mu_{wk}} = \frac{q_{dk} T_{dw}}{\lambda}$

$\because \sum_{w,k} \mu_{wk} = \frac{\sum_{w,k} \sum_d q_{d,k} T_{dw}}{\lambda}$

$\Rightarrow 1 = \frac{d}{\lambda}$

$\Rightarrow \lambda = d$

$\therefore \hat{\mu_{wk}} = \frac{q_{wk} T_{dw}}{d},\ \forall q_{dk} = \frac{p(d|C_d=k)p(C_d=k)}{p(d)} = \frac{\frac{n_d!}{\prod_w T_{dw}!} \pi_k \prod_w \mu_{wk}^{T_{dw}}}{\frac{n_d!}{\prod_w T_{dw}!} \sum_k \pi_k \prod_w \mu_{wk}^{T_{dw}}}$

$\because \sum_k \pi_k = 1$

$\Rightarrow By\ Lagrange\ multiplier,\ B^{'}(\pi) = B(\pi) + \lambda(1 - \sum_k \pi_k)$

$\Rightarrow \frac{dB^{'}(\pi)}{d\pi_k} = \frac{dB(\pi)}{d\pi_k} - \lambda = 0$

$\Rightarrow \frac{d}{d\pi_k} \sum_d \sum_k q_{dk} \ln \pi_k = \lambda$

$= \frac{1}{\pi_k} \sum_d q_{dk} = \lambda$

$\Rightarrow \hat{\pi_k} = \frac{\sum_d q_{dk}}{\lambda} = \frac{\sum_d q_{dk}}{d}$

因此只要以(1)=E和(2)=M不段迭代即可实现EM for mixture of multinomials

Problem 5

Implement the EM algorithm

1. 此题想问在topic `k=10,20,30,50` 时，每个topic中出现最频繁的words。
   首先这笔数据需要先将 `libsvm` 的 `[id:count]` 做处理以形成T_{dw} ，再应用EM算法返回$\mu_{wk}, \pi k$。
   如果想知道每个topic出现最频繁的words可以在μ_{wk} 寻找每个 `k` 所对应的w最大值，并匹配 `vocab` 的 `id` , `word` ，以此来寻找每个topic中出现最频繁的words，可以以此来判断不同topic `k` 所对应的主题：

1. 我认为在 `k=20, 30` 时表现最好，在这个应用里并不是k越大越好，而是k最合适最好。这是一个非监督式学习的应用，EM的延伸应用就是k means，因为我们不知道category数量有多少，因此设定太多太少都不合适。此题也是如此，由于原本不知道k的数量，因此过大过小都不合适，然而此题是来分类topic，而不是像平时我们看k means，直接画出数据分类结果的图即可，在此题可以用判断的方式来检查每个分类结果是否优良，像 `k=10` 比较难以看出哪个topic应该总结成什么，而 `k=50` 则又分类太细，当然还是依照自己的目的以及迭代收敛速度为主来判断最好。

- k=10

```
Topic 0:
    jesus: 0.002355
    israel: 0.002244
    years: 0.002212
    space: 0.001971
    point: 0.001894
    things: 0.001808
    government: 0.001726
    jews: 0.001700
    true: 0.001628
    university: 0.001589
Topic 1:
    going: 0.002309
    really: 0.002206
    didnt: 0.002204
    car: 0.001864
    things: 0.001831
    come: 0.001802
    got: 0.001678
    government: 0.001644
    look: 0.001598
    believe: 0.001545
Topic 2:
    drive: 0.005186
    scsi: 0.002892
    problem: 0.002698
    disk: 0.002674
    hard: 0.002552
    believe: 0.002193
    drives: 0.002130
    years: 0.002108
    really: 0.002086
    card: 0.002048
Topic 3:
    available: 0.002653
    file: 0.002615
    information: 0.002367
    send: 0.002198
    software: 0.002173
    files: 0.002160
    key: 0.002131
    computer: 0.002110
    data: 0.002063
    thanks: 0.002037
Topic 4:
    armenian: 0.002566
    team: 0.002342
    turkish: 0.002310
    year: 0.002228
    years: 0.002217
    armenians: 0.002017
    turkey: 0.001837
    game: 0.001825
    dos: 0.001725
    window: 0.001675
Topic 5:
    believe: 0.002631
    lord: 0.002530
    game: 0.002254
    drive: 0.002189
    jehovah: 0.002168
    christ: 0.002027
    jesus: 0.002027
    going: 0.001903
    doesnt: 0.001893
    things: 0.001882
Topic 6:
    file: 0.005101
    number: 0.002774
    law: 0.002714
    program: 0.002263
    year: 0.002241
    entry: 0.002159
    jesus: 0.002142
    believe: 0.001902
    really: 0.001809
    information: 0.001794
Topic 7:
    jpeg: 0.005165
    file: 0.003541
    image: 0.003470
    information: 0.002697
    available: 0.002519
    files: 0.002492
    gif: 0.002351
    government: 0.002351
    version: 0.002249
    software: 0.002084
Topic 8:
    image: 0.003168
    stephanopoulos: 0.003073
    using: 0.002654
    program: 0.002604
    file: 0.002583
    going: 0.002348
    available: 0.002313
    space: 0.002302
    information: 0.002257
    problem: 0.002225
Topic 9:
    windows: 0.004624
    thanks: 0.002943
    problem: 0.002830
    using: 0.002448
    card: 0.002315
    help: 0.002202
    game: 0.002201
    dos: 0.002162
    government: 0.001980
    version: 0.001965
```

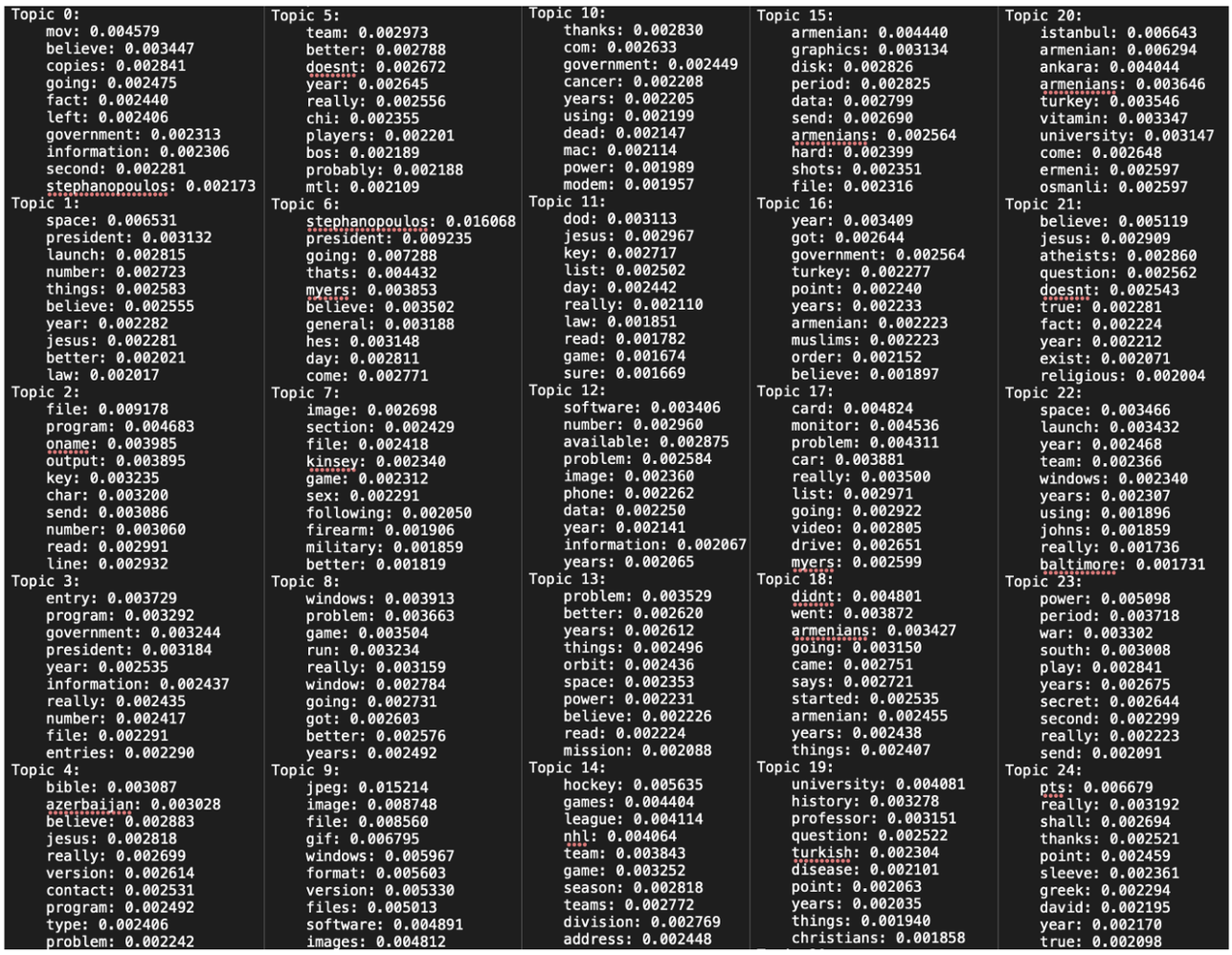
- k=20

```
Topic 0:
    armenian: 0.006583
    armenians: 0.005221
    didnt: 0.004687
    turkish: 0.003708
    went: 0.002975
    genocide: 0.002926
    going: 0.002810
    come: 0.002603
    government: 0.002573
    available: 0.002525
Topic 1:
    jesus: 0.003632
    game: 0.002537
    really: 0.002512
    believe: 0.002367
    things: 0.002218
    point: 0.002180
    didnt: 0.002137
    look: 0.002060
    law: 0.002051
    thats: 0.002019
Topic 2:
    window: 0.002717
    space: 0.002615
    using: 0.002164
    program: 0.002043
    thanks: 0.001892
    question: 0.001854
    server: 0.001849
    information: 0.001835
    help: 0.001795
    available: 0.001766
Topic 3:
    file: 0.008584
    entry: 0.003668
    program: 0.002916
    subject: 0.002383
    entries: 0.002300
    space: 0.002299
    available: 0.002183
    gun: 0.001963
    using: 0.001913
    number: 0.001910
Topic 4:
    mhz: 0.003330
    car: 0.002480
    true: 0.002331
    going: 0.002327
    thanks: 0.002325
    windows: 0.002178
    years: 0.002037
    chip: 0.001964
    using: 0.001953
    problem: 0.001947
Topic 5:
    jpeg: 0.011722
    file: 0.007494
    image: 0.006520
    gif: 0.005392
    dos: 0.004658
    windows: 0.004482
    files: 0.004413
    version: 0.004143
    images: 0.004034
    format: 0.003975
Topic 6:
    drive: 0.004427
    scsi: 0.004021
    problem: 0.003137
    mac: 0.002735
    windows: 0.002361
    really: 0.002355
    hard: 0.002196
    things: 0.002187
    jehovah: 0.002174
    using: 0.002125
Topic 7:
    water: 0.002580
    image: 0.002324
    president: 0.002166
    government: 0.002129
    armenian: 0.002108
    years: 0.001930
    look: 0.001924
    problem: 0.001877
    going: 0.001869
    things: 0.001844
Topic 8:
    encryption: 0.003082
    appears: 0.002572
    government: 0.002510
    chip: 0.002387
    art: 0.002312
    widget: 0.002046
    data: 0.002031
    using: 0.002020
    law: 0.001964
    technology: 0.001930
Topic 9:
    image: 0.004598
    thanks: 0.003136
    software: 0.002944
    graphics: 0.002837
    data: 0.002769
    program: 0.002745
    mail: 0.002697
    using: 0.002684
    send: 0.002680
    file: 0.002648
Topic 10:
    file: 0.003736
    game: 0.002653
    year: 0.002533
    program: 0.002437
    team: 0.002394
    space: 0.002321
    games: 0.002286
    information: 0.002115
    hockey: 0.002114
    output: 0.001824
Topic 11:
    stephanopoulos: 0.002935
    lost: 0.002469
    won: 0.002225
    data: 0.002148
    space: 0.002131
    launch: 0.002100
    going: 0.002047
    probably: 0.001925
    year: 0.001917
    key: 0.001884
Topic 12:
    information: 0.002455
    number: 0.002320
    problem: 0.002079
    government: 0.002063
    help: 0.001978
    read: 0.001960
    believe: 0.001886
    wire: 0.001808
    thanks: 0.001792
    going: 0.001766
Topic 13:
    key: 0.004845
    chip: 0.003143
    number: 0.002577
    problem: 0.002445
    got: 0.002240
    law: 0.002162
    really: 0.002091
    believe: 0.001950
    game: 0.001928
    doesnt: 0.001786
Topic 14:
    believe: 0.003559
    point: 0.002171
    power: 0.002091
    problem: 0.001994
    windows: 0.001921
    really: 0.001914
    fact: 0.001827
    got: 0.001805
    play: 0.001791
    religion: 0.001754
Topic 15:
    information: 0.002852
    going: 0.002624
    drive: 0.002620
    space: 0.002601
    software: 0.002292
    list: 0.002224
    data: 0.002202
    available: 0.002154
    president: 0.002006
    stephanopoulos: 0.001904
Topic 16:
    data: 0.003420
    image: 0.002171
    government: 0.002140
    available: 0.002125
    better: 0.002117
    really: 0.002007
    point: 0.001865
    believe: 0.001751
    years: 0.001727
    going: 0.001707
Topic 17:
    adl: 0.002650
    state: 0.002573
    government: 0.002487
    information: 0.002398
    number: 0.002008
    law: 0.001977
    kuwait: 0.001873
    general: 0.001794
    public: 0.001745
    president: 0.001715
Topic 18:
    jesus: 0.003586
    theory: 0.002567
    universe: 0.002501
    really: 0.002421
    better: 0.002410
    believe: 0.002396
    life: 0.002384
    hell: 0.002322
    evidence: 0.002273
    years: 0.002259
Topic 19:
    drive: 0.004126
    card: 0.002582
    disk: 0.002578
    thanks: 0.002266
    really: 0.002251
    hard: 0.002132
    pts: 0.002113
    using: 0.001995
    game: 0.001988
    jews: 0.001955
```

- k=30

```
Topic 0:
    jpeg: 0.012581
    file: 0.007392
    image: 0.006501
    myers: 0.006172
    gif: 0.004952
    files: 0.004366
    available: 0.003847
    images: 0.003785
    format: 0.003744
    president: 0.003601
Topic 1:
    drive: 0.003758
    mov: 0.003370
    power: 0.002884
    monitor: 0.002618
    stephanopoulos: 0.002583
    card: 0.002538
    bit: 0.002475
    thanks: 0.002369
    going: 0.002212
    best: 0.002199
Topic 2:
    problem: 0.003910
    using: 0.003619
    available: 0.002967
    program: 0.002805
    thanks: 0.002702
    information: 0.002529
    help: 0.002312
    space: 0.002280
    disk: 0.002211
    ftp: 0.002109
Topic 3:
    car: 0.002751
    really: 0.002728
    question: 0.002693
    years: 0.002556
    windows: 0.002437
    thanks: 0.002412
    des: 0.002322
    data: 0.002292
    power: 0.002119
    key: 0.002005
Topic 4:
    better: 0.002850
    church: 0.002516
    true: 0.002453
    believe: 0.002383
    jesus: 0.002337
    tell: 0.002326
    really: 0.002314
    place: 0.002164
    law: 0.002090
    man: 0.002083
Topic 5:
    game: 0.004907
    problem: 0.003190
    drive: 0.003175
    better: 0.002829
    really: 0.002529
    got: 0.002527
    windows: 0.002416
    team: 0.002231
    year: 0.002185
    years: 0.002150
Topic 6:
    windows: 0.005922
    card: 0.004094
    dos: 0.003923
    key: 0.003790
    problem: 0.003221
    disk: 0.003205
    chip: 0.003200
    using: 0.002769
    mouse: 0.002706
    drive: 0.002515
Topic 7:
    appears: 0.003951
    widget: 0.003213
    art: 0.003182
    list: 0.003160
    set: 0.003088
    information: 0.002653
    data: 0.002499
    file: 0.002480
    application: 0.002239
    code: 0.002004
Topic 8:
    wire: 0.002832
    thanks: 0.002409
    wiring: 0.002333
    power: 0.002229
    information: 0.002198
    ground: 0.002129
    believe: 0.002075
    come: 0.002071
    point: 0.002021
    subject: 0.001984
Topic 9:
    dos: 0.002726
    number: 0.002554
    problem: 0.002434
    point: 0.002178
    data: 0.002099
    question: 0.001861
    gun: 0.001857
    doesnt: 0.001820
    going: 0.001798
    using: 0.001787
Topic 10:
    windows: 0.004395
    file: 0.003493
    law: 0.002804
    problem: 0.002802
    true: 0.002664
    files: 0.002621
    argument: 0.002569
    believe: 0.002490
    really: 0.002461
    using: 0.002446
Topic 11:
    image: 0.008021
    data: 0.005954
    available: 0.004426
    software: 0.004216
    graphics: 0.003315
    using: 0.002909
    systems: 0.002877
    ftp: 0.002735
    file: 0.002540
    processing: 0.002466
Topic 12:
    really: 0.003714
    year: 0.002736
    jesus: 0.002568
    thanks: 0.002502
    engine: 0.002425
    got: 0.002349
    problem: 0.002325
    look: 0.002231
    car: 0.002204
    things: 0.002174
Topic 13:
    jehovah: 0.003299
    stephanopoulos: 0.003247
    going: 0.002951
    lord: 0.002925
    year: 0.002602
    believe: 0.002537
    elohim: 0.002512
    thats: 0.002167
    years: 0.002140
    better: 0.002120
Topic 14:
    bos: 0.003106
    window: 0.002901
    chi: 0.002832
    tor: 0.002809
    det: 0.002786
    available: 0.002773
    nyi: 0.002535
    que: 0.002512
    van: 0.002467
    buf: 0.002467
Topic 15:
    hockey: 0.003335
    games: 0.002911
    team: 0.002619
    information: 0.002514
    league: 0.002324
    believe: 0.002229
    internet: 0.002025
    world: 0.002018
    nhl: 0.001934
    belief: 0.001921
Topic 16:
    period: 0.003418
    play: 0.002756
    second: 0.002704
    season: 0.002603
    believe: 0.002557
    team: 0.002409
    soon: 0.002246
    year: 0.002147
    key: 0.002136
    best: 0.002068
Topic 17:
    going: 0.003544
    stephanopoulos: 0.003310
    things: 0.002770
    really: 0.002701
    didnt: 0.002504
    car: 0.002341
    thats: 0.002289
    says: 0.002261
    fact: 0.002032
    youre: 0.002022
Topic 18:
    bit: 0.003126
    didnt: 0.002771
    windows: 0.002738
    dod: 0.002559
    come: 0.002431
    things: 0.002333
    available: 0.002324
    program: 0.002310
    help: 0.002297
    going: 0.002243
Topic 19:
    government: 0.003029
    turkish: 0.002830
    years: 0.002348
    believe: 0.002195
    jews: 0.001762
    really: 0.001731
    point: 0.001655
    university: 0.001640
    send: 0.001635
    year: 0.001632
Topic 20:
    believe: 0.002702
    really: 0.002513
    drive: 0.002480
    doesnt: 0.002384
    thats: 0.002241
    key: 0.002129
    going: 0.002082
    sure: 0.002034
    number: 0.001975
    point: 0.001935
Topic 21:
    image: 0.002977
    information: 0.002953
    list: 0.002771
    university: 0.002613
    available: 0.002435
    file: 0.002402
    data: 0.002241
    thanks: 0.002226
    really: 0.002028
    question: 0.001995
Topic 22:
    information: 0.003546
    available: 0.003330
    server: 0.002666
    file: 0.002525
    software: 0.002500
    computer: 0.002414
    university: 0.002292
    version: 0.002283
    sun: 0.002250
    windows: 0.002184
Topic 23:
    drive: 0.009499
    drives: 0.006188
    disk: 0.005847
    hard: 0.005371
    controller: 0.004256
    help: 0.003239
    bios: 0.003211
    type: 0.003045
    version: 0.003023
    thanks: 0.003015
Topic 24:
    car: 0.003803
    window: 0.002929
    using: 0.002369
    thats: 0.002246
    sure: 0.002233
    really: 0.002174
    problem: 0.002101
    better: 0.002052
    believe: 0.002050
    years: 0.001898
Topic 25:
    space: 0.003203
    problem: 0.002763
    world: 0.002704
    war: 0.002666
    send: 0.002484
    data: 0.002436
    south: 0.002180
    using: 0.002112
    years: 0.002012
    secret: 0.001886
Topic 26:
    file: 0.013644
    program: 0.006001
    entry: 0.005223
    launch: 0.003288
    space: 0.003270
    read: 0.003164
    number: 0.002855
    year: 0.002813
    output: 0.002689
    send: 0.002609
Topic 27:
    jesus: 0.003237
    things: 0.002797
    thing: 0.002518
    children: 0.002514
    game: 0.002296
    question: 0.002126
    says: 0.002042
    didnt: 0.002011
    going: 0.001932
    really: 0.001926
Topic 28:
    jesus: 0.003290
    believe: 0.002280
    problem: 0.002155
    really: 0.002123
    government: 0.002056
    year: 0.001963
    come: 0.001891
    years: 0.001852
    law: 0.001829
    information: 0.001744
Topic 29:
    armenian: 0.006233
    armenians: 0.003852
    space: 0.003396
    pts: 0.002568
    russian: 0.002443
    adl: 0.002381
    years: 0.002340
    turkish: 0.002297
    venus: 0.002236
    turks: 0.002029
```

- k=50


Topic 0:
mov: 0.004579
believe: 0.003447
copies: 0.002841
going: 0.002475
fact: 0.002440
left: 0.002406
government: 0.002313
information: 0.002306
second: 0.002281
stephanopoulos: 0.002173
Topic 1:
space: 0.006531
president: 0.003132
launch: 0.002815
number: 0.002723
things: 0.002583
believe: 0.002555
year: 0.002282
jesus: 0.002281
better: 0.002021
law: 0.002017
Topic 2:
file: 0.009178
program: 0.004683
oname: 0.003985
output: 0.003895
key: 0.003235
char: 0.003200
send: 0.003086
number: 0.003060
read: 0.002991
line: 0.002932
Topic 3:
entry: 0.003729
program: 0.003292
government: 0.003244
president: 0.003184
year: 0.002535
information: 0.002437
really: 0.002435
number: 0.002417
file: 0.002291
entries: 0.002290
Topic 4:
bible: 0.003087
azerbaijan: 0.003028
believe: 0.002883
jesus: 0.002818
really: 0.002699
version: 0.002614
contact: 0.002531
program: 0.002492
type: 0.002406
problem: 0.002242
Topic 5:
team: 0.002973
better: 0.002788
doesnt: 0.002672
year: 0.002645
really: 0.002556
chi: 0.002355
players: 0.002201
bos: 0.002189
probably: 0.002188
mtl: 0.002109
Topic 6:
stephanopoulos: 0.016068
president: 0.009235
going: 0.007288
thats: 0.004432
myers: 0.003853
believe: 0.003502
general: 0.003188
hes: 0.003148
day: 0.002811
come: 0.002771
Topic 7:
image: 0.002698
section: 0.002429
file: 0.002418
kinsey: 0.002340
game: 0.002312
sex: 0.002291
following: 0.002050
firearm: 0.001906
military: 0.001859
better: 0.001819
Topic 8:
windows: 0.003913
problem: 0.003663
game: 0.003504
run: 0.003234
really: 0.003159
window: 0.002784
going: 0.002731
got: 0.002603
better: 0.002576
years: 0.002492
Topic 9:
jpeg: 0.015214
image: 0.008748
file: 0.008560
gif: 0.006795
windows: 0.005967
format: 0.005603
version: 0.005330
files: 0.005013
software: 0.004891
images: 0.004812
Topic 10:
thanks: 0.002830
com: 0.002633
government: 0.002449
cancer: 0.002208
years: 0.002205
using: 0.002199
dead: 0.002147
mac: 0.002114
power: 0.001989
modem: 0.001957
Topic 11:
dod: 0.003113
jesus: 0.002967
key: 0.002717
list: 0.002502
day: 0.002442
really: 0.002110
law: 0.001851
read: 0.001782
game: 0.001674
sure: 0.001669
Topic 12:
software: 0.003406
number: 0.002960
available: 0.002875
problem: 0.002584
image: 0.002360
phone: 0.002262
data: 0.002250
year: 0.002141
information: 0.002067
years: 0.002065
Topic 13:
problem: 0.003529
better: 0.002620
years: 0.002612
things: 0.002496
orbit: 0.002436
space: 0.002353
power: 0.002231
believe: 0.002226
read: 0.002224
mission: 0.002088
Topic 14:
hockey: 0.005635
games: 0.004404
league: 0.004114
nhl: 0.004064
team: 0.003843
game: 0.003252
season: 0.002818
teams: 0.002772
division: 0.002769
address: 0.002448
Topic 15:
armenian: 0.004440
graphics: 0.003134
disk: 0.002826
period: 0.002825
data: 0.002799
send: 0.002690
armenians: 0.002564
hard: 0.002399
shots: 0.002351
file: 0.002316
Topic 16:
year: 0.003409
got: 0.002644
government: 0.002564
turkey: 0.002277
point: 0.002240
years: 0.002233
armenian: 0.002223
muslims: 0.002223
order: 0.002152
believe: 0.001897
Topic 17:
card: 0.004824
monitor: 0.004536
problem: 0.004311
car: 0.003881
really: 0.003500
list: 0.002971
going: 0.002922
video: 0.002805
drive: 0.002651
myers: 0.002599
Topic 18:
didnt: 0.004801
went: 0.003872
armenians: 0.003427
going: 0.003150
came: 0.002751
says: 0.002721
started: 0.002535
armenian: 0.002455
years: 0.002438
things: 0.002407
Topic 19:
university: 0.004081
history: 0.003278
professor: 0.003151
question: 0.002522
turkish: 0.002304
disease: 0.002101
point: 0.002063
years: 0.002035
things: 0.001940
christians: 0.001858
Topic 20:
istanbul: 0.006643
armenian: 0.006294
ankara: 0.004044
armenians: 0.003646
turkey: 0.003546
vitamin: 0.003347
university: 0.003147
come: 0.002648
ermeni: 0.002597
osmanli: 0.002597
Topic 21:
believe: 0.005119
jesus: 0.002909
atheists: 0.002860
question: 0.002562
doesnt: 0.002543
true: 0.002281
fact: 0.002224
year: 0.002212
exist: 0.002071
religious: 0.002004
Topic 22:
space: 0.003466
launch: 0.003432
year: 0.002468
team: 0.002366
windows: 0.002340
years: 0.002307
using: 0.001896
johns: 0.001859
really: 0.001736
baltimore: 0.001731
Topic 23:
power: 0.005098
period: 0.003718
war: 0.003302
south: 0.003008
play: 0.002841
years: 0.002675
secret: 0.002644
second: 0.002299
really: 0.002223
send: 0.002091
Topic 24:
pts: 0.006679
really: 0.003192
shall: 0.002694
thanks: 0.002521
point: 0.002459
sleeve: 0.002361
greek: 0.002294
david: 0.002195
year: 0.002170
true: 0.002098


Topic 25:
key: 0.003422
game: 0.002597
program: 0.002475
drug: 0.002475
information: 0.002392
number: 0.002266
year: 0.002179
security: 0.002171
phigs: 0.002131
public: 0.002096
Topic 26:
flyers: 0.004934
problem: 0.003542
game: 0.003433
puck: 0.003232
best: 0.003194
got: 0.003128
play: 0.003102
better: 0.003076
power: 0.002665
lot: 0.002558
Topic 27:
ground: 0.003830
drug: 0.003141
government: 0.002441
states: 0.002223
president: 0.002111
great: 0.002058
station: 0.002003
book: 0.001905
children: 0.001787
war: 0.001781
Topic 28:
image: 0.005358
data: 0.004360
available: 0.004202
lord: 0.002463
program: 0.002452
sun: 0.002417
information: 0.002416
software: 0.002328
ftp: 0.002265
jesus: 0.002214
Topic 29:
drive: 0.004690
problem: 0.004284
tape: 0.003557
problems: 0.003071
son: 0.002986
believe: 0.002213
using: 0.002209
disk: 0.002089
read: 0.002079
hard: 0.001981
Topic 30:
encryption: 0.005921
chip: 0.005168
government: 0.004329
technology: 0.003579
card: 0.003380
data: 0.002912
information: 0.002844
clipper: 0.002696
thanks: 0.002623
access: 0.002585
Topic 31:
game: 0.002972
question: 0.002328
wrong: 0.002318
team: 0.002218
law: 0.002126
year: 0.002078
look: 0.002056
different: 0.002035
paul: 0.001929
man: 0.001881
Topic 32:
israel: 0.003890
list: 0.003570
key: 0.003090
allocation: 0.003049
program: 0.002815
using: 0.002705
unit: 0.002697
government: 0.002642
problem: 0.002535
cross: 0.002404
Topic 33:
file: 0.009566
gun: 0.003568
information: 0.003248
law: 0.002750
control: 0.002370
believe: 0.002294
police: 0.002260
states: 0.002222
privacy: 0.002077
internet: 0.002056
Topic 34:
myers: 0.004349
going: 0.003211
mhz: 0.003148
government: 0.002841
years: 0.002691
church: 0.002681
game: 0.002531
pope: 0.002248
far: 0.002219
case: 0.002176
Topic 35:
thanks: 0.003037
university: 0.002821
van: 0.002460
problem: 0.002316
using: 0.002288
history: 0.002244
doug: 0.002243
going: 0.002175
information: 0.002171
jews: 0.002025
Topic 36:
file: 0.005682
thanks: 0.005534
windows: 0.004917
server: 0.004013
help: 0.004010
using: 0.003937
available: 0.003167
version: 0.003031
software: 0.002914
running: 0.002867
Topic 37:
lost: 0.007633
won: 0.006796
problem: 0.002972
jews: 0.002904
idle: 0.002904
thanks: 0.002504
national: 0.002410
computer: 0.002299
american: 0.002286
chicago: 0.002231
Topic 38:
adl: 0.003917
ripem: 0.003333
key: 0.002749
believe: 0.002626
game: 0.002481
problem: 0.002341
drive: 0.002201
rsa: 0.002130
using: 0.002098
point: 0.002072
Topic 39:
government: 0.004789
drive: 0.004435
israel: 0.004159
turkish: 0.003962
rights: 0.003083
law: 0.003041
master: 0.002979
muslim: 0.002935
armenians: 0.002739
slave: 0.002735
Topic 40:
windows: 0.006131
thanks: 0.004934
openwindows: 0.004886
window: 0.004131
file: 0.004056
dos: 0.003824
run: 0.003596
program: 0.003543
version: 0.003325
using: 0.003253
Topic 41:
believe: 0.003468
ndetloopc: 0.003261
evidence: 0.002849
windows: 0.002841
world: 0.002719
using: 0.002687
christians: 0.002485
dos: 0.002478
really: 0.002361
point: 0.002357
Topic 42:
key: 0.004489
window: 0.004305
kuwait: 0.004252
using: 0.003122
program: 0.002469
second: 0.002284
bit: 0.002282
keys: 0.002281
image: 0.002246
chip: 0.002242
Topic 43:
question: 0.002926
drive: 0.002638
key: 0.002356
point: 0.002239
pain: 0.002231
going: 0.002157
public: 0.002144
course: 0.002025
second: 0.002013
thats: 0.001964
Topic 44:
life: 0.003187
christ: 0.003018
jesus: 0.002990
space: 0.002896
believe: 0.002157
point: 0.002059
word: 0.001986
bible: 0.001978
john: 0.001946
world: 0.001930
Topic 45:
dos: 0.007499
windows: 0.004658
keyboard: 0.003773
information: 0.003679
thanks: 0.003400
list: 0.003215
cancer: 0.003156
available: 0.003070
number: 0.002710
using: 0.002681
Topic 46:
appears: 0.005130
art: 0.004331
software: 0.003196
disk: 0.002992
mac: 0.002890
files: 0.002840
wolverine: 0.002493
problem: 0.002424
argument: 0.002389
hiv: 0.002213
Topic 47:
pts: 0.003051
power: 0.002636
card: 0.002601
list: 0.002314
information: 0.002287
long: 0.002276
point: 0.002229
sure: 0.002228
drive: 0.002225
problem: 0.002220
Topic 48:
jews: 0.003468
church: 0.003269
turkish: 0.003144
widget: 0.002961
believe: 0.002960
water: 0.002819
data: 0.002690
program: 0.002635
going: 0.002527
file: 0.002496
Topic 49:
drive: 0.008336
scsi: 0.005918
ide: 0.004684
problem: 0.003460
believe: 0.003272
hard: 0.003149
controller: 0.003052
drives: 0.003050
thanks: 0.002878
windows: 0.002838