Spamerkennung mit log-linearen Modellen

Implementieren Sie ein weiteres Spamerkennungs-System auf Basis von log-linearen Modellen mit L_2-Regularisierung, welches dieselben Daten wie in der letzten Übung verwendet. Es sollte wieder aus einem Trainingsprogramm und einem Anwendungsprogramm bestehen.

Hier ist Pseudocode für die Aufgabe:

```
for n epochs
    for mail in data
        p(class|mail) für alle Klassen class berechnen
        beobachtete Merkmalswerte berechnen
        erwartete Merkmalswerte berechnen
        Gradient berechnen
        Gewichtsvektor anpassen
```

Die Programme sollen folgendermaßen aufgerufen werden:

```
python3 train.py train-dir paramfile
```

```
python3 test.py paramfile mail-dir
```

Vorüberlegungen

- Welche Merkmale verwenden Sie am besten?
- Welche Teilaufgaben umfasst das Training?
- Welche Datenstrukturen verwenden Sie?
- Was speichern Sie in der Parameterdatei?
- Was ändert sich durch die L_2-Regularisierung?
- Welche Teilaufgaben umfasst das Anwendungsprogramm?

Schicken Sie das fertige **Programm**, die optimierten Werte für **Lernrate** und **Regularisierung** sowie die Liste der für die Testdaten **ausgegebenen Klassen** an `schmid@cis.lmu.de`.