**Logistic regression:**

Penalty: L1

C=1.0 Inverse of regularization strength.

Classification rate: 89.74%.

Noun phrases with a percentage greater than 50% are regional noun phrases.

Number of regional noun phrases: 65.

**Logistic regression:**

Penalty: L2

C=1.0 Inverse of regularization strength.

Classification rate: 89.74%.

Noun phrases with a percentage greater than 50% are regional noun phrases.

Number of regional noun phrases: 67 (‘inch meat feast’, ‘rubicon mango’).

**L2 logistic (Multinomial) regression:**

Penalty: L2

C=1.0 Inverse of regularization strength.

Classification rate: 61.53%.

This situation cannot be indicated that a phrase is regional, because the regional probability of all phrases is around 50%.

**Linear SVC:**

C=1.0 Inverse of regularization strength.

Classification rate: 79.48 %.

This situation cannot be indicated that a phrase which is larger than a certain probability is regional as many regional phrases such as ‘bru’, ‘mince’ and ‘vegetable pakora’ are determined as non-regional. At the same time, many non-regional phrases such as ‘email sms’, ‘half’ and ‘everything’ have more than 50% probability. As a consequence, regional phrases cannot be correctly classified by using Linear SVC.

**Conclusion**

By comparing four classifiers, logistic regression is the most suitable classifier for this project. Furthermore, the project will use L2 as the penalty parameter.

**Logistic regression:**

Logistic regression is well suited to describe the relationship which is expressed as probability between classification results and one or more classifications [5]. It can adapt to multiple classification results. In this project, logistic regression is used to calculate the probability of a binary event occurring under multiple independent features [4].

Model:

Cost function:

Problem:

Overfitting problem: We consider supervised learning in settings where there are many input features, but where there is a small subset of the features that is sufficient to approximate the target concept well. In supervised learning settings with many input features, overfitting is usually a potential problem unless there is ample training data

L1 and L2: Lasso (l1-)penalties are useful for fitting a wide variety of models. Newly developed computational algorithms allow application of these models to large data sets, exploiting sparsity for both statistical and computation gains[1]

Comparsison: Applying an L2 penalty tends to result in all small but non-zero regression coefficients, whereas applying an L1 penalty tends to result in many regression coefficients shrunk exactly to zero and a few other regression coefficients with comparatively little shrinkage[2]

L1 regularization, uses a penalty term which encourages the sum of the absolute values of the parameters to be small. The second, L2 regularization, encourages the sum of the squares of the parameters to be small. It has frequently been observed that L1 regularization in many models causes many parameters to equal zero, so that the parameter vector is sparse. This makes it a natural candidate in feature selection settings, where we believe that many features should be ignored.[3]

**References**

[1] Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273-282.

[2] Goeman, J., Meijer, R., & Chaturvedi, N. (2012). L1 and L2 penalized regression models. *cran. r-project. or*.

[3] Ng, A. Y. (2004, July). Feature selection, L 1 vs. L 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning* (p. 78). ACM.

[4] Walker, S. H., & Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, *54*(1-2), 167-179.