**COMP 551: Applied Machine Learning**

**Assignment #2**

**Report**

By:

Weishi Wang (260540022)

Submitted to:

Prof. Sarath Chandar

Oct 21st, 2018

**Question 1.**

Both IMDB and yelp data are converted into BBoW and FBoW forms. See function “modify”,“BBoW” and “FBoW".

**Question 2.**

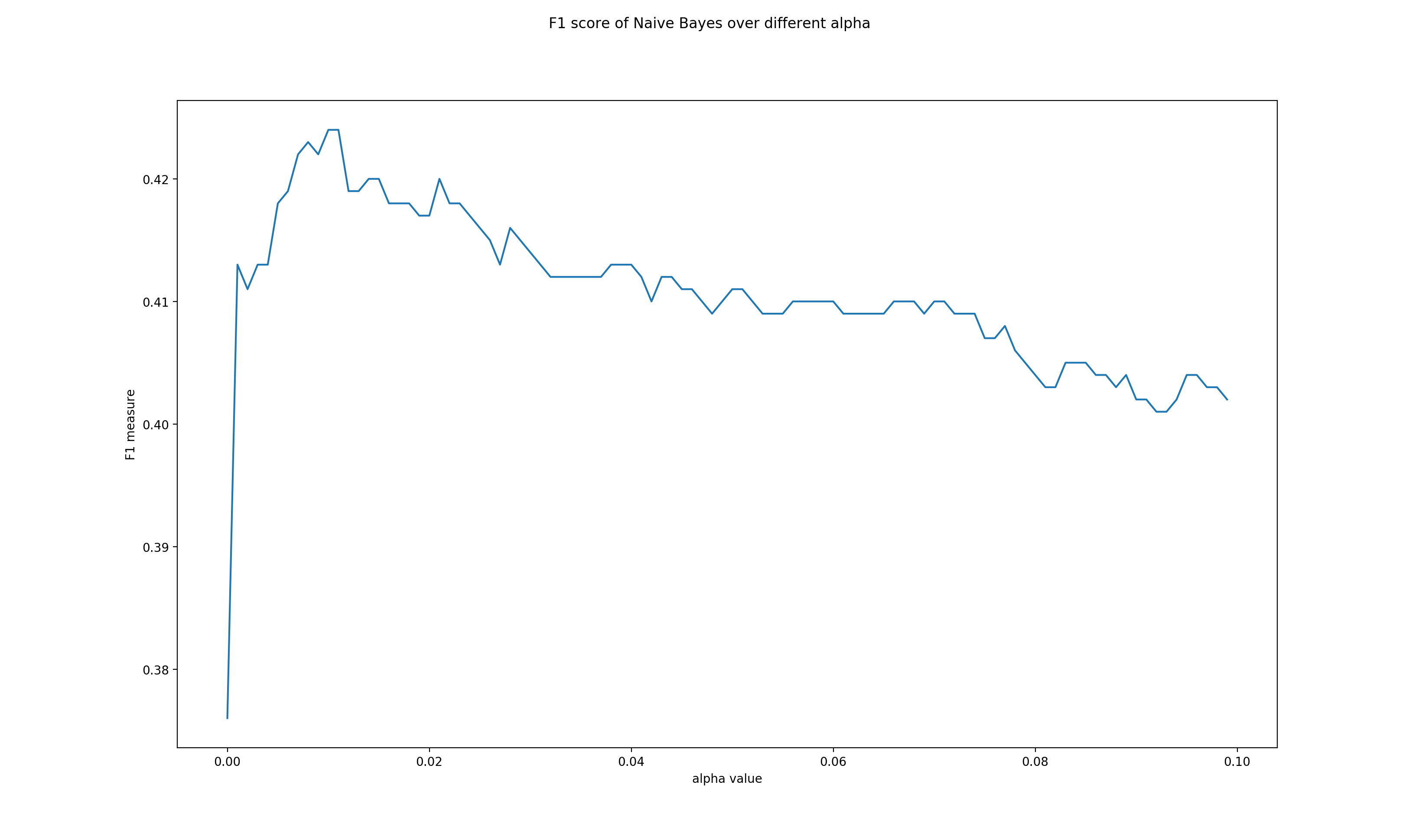
**(a)** The performance of random uniform is: 0.1885

The performance of random majority is: 0.351

**(b)**

**1.** Tune Naïve Bayes:

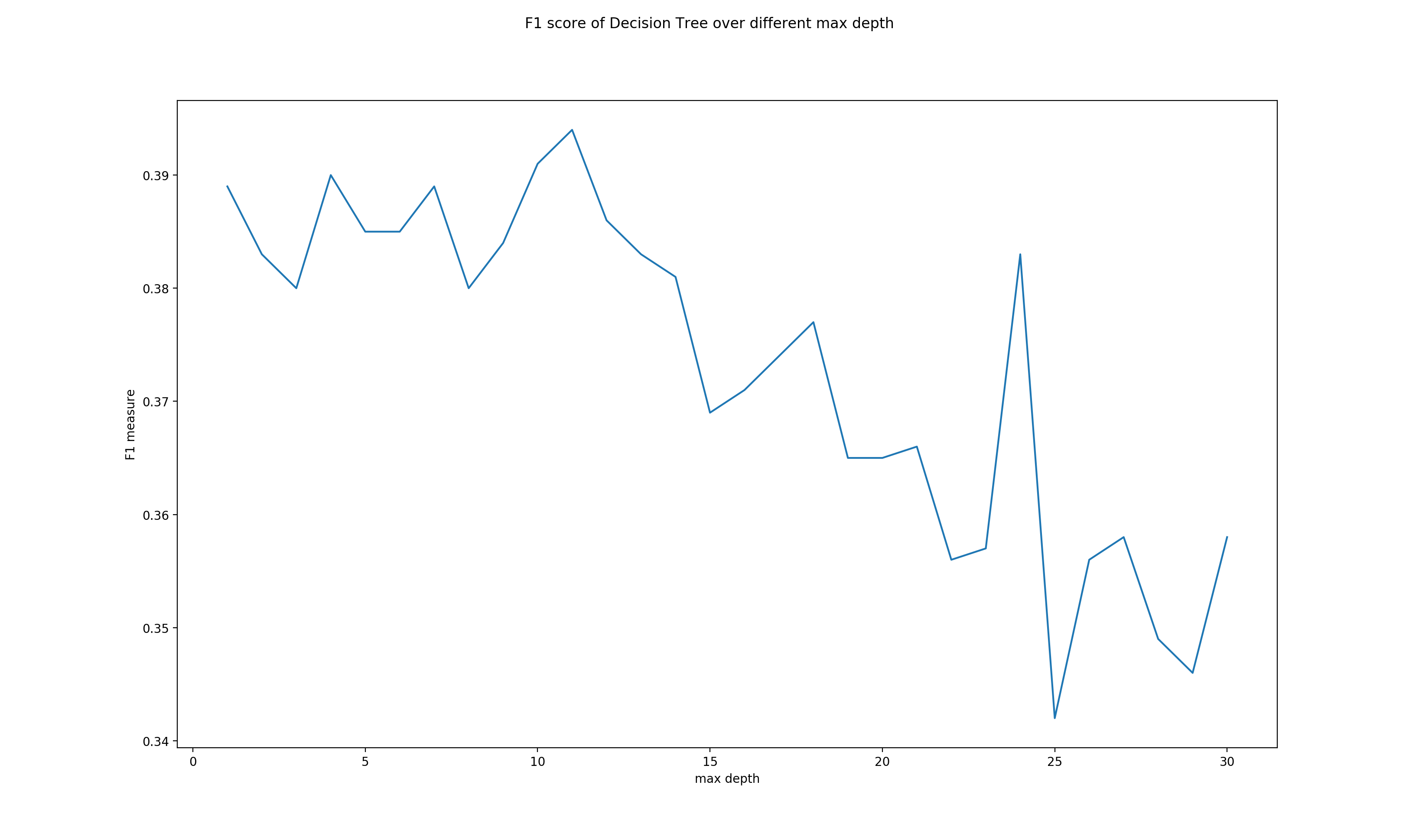
alpha:



best alpha: 0.0100 (range: 1e-10 to 0.1)

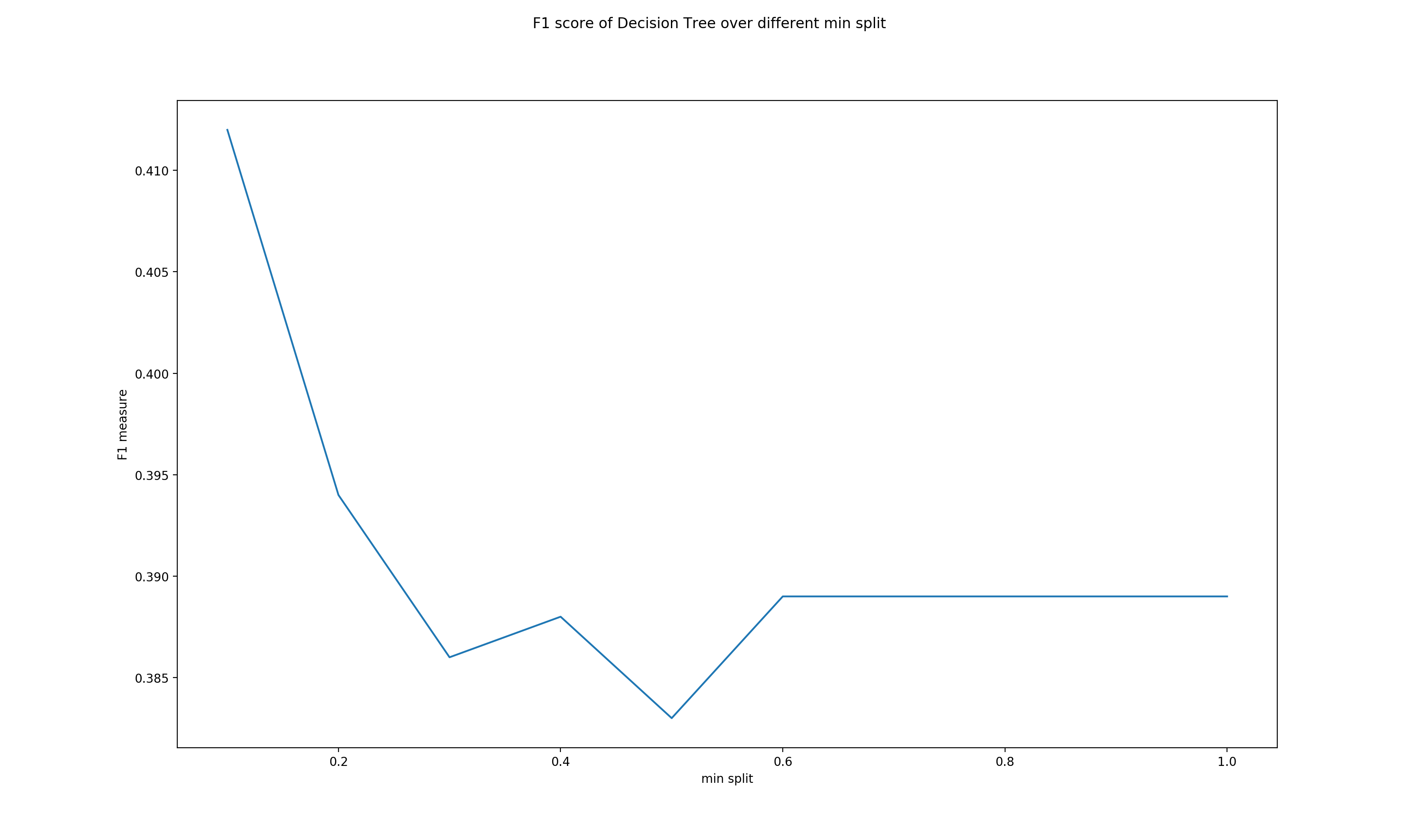
**2.** Tune Decision Tree:

**depth:**



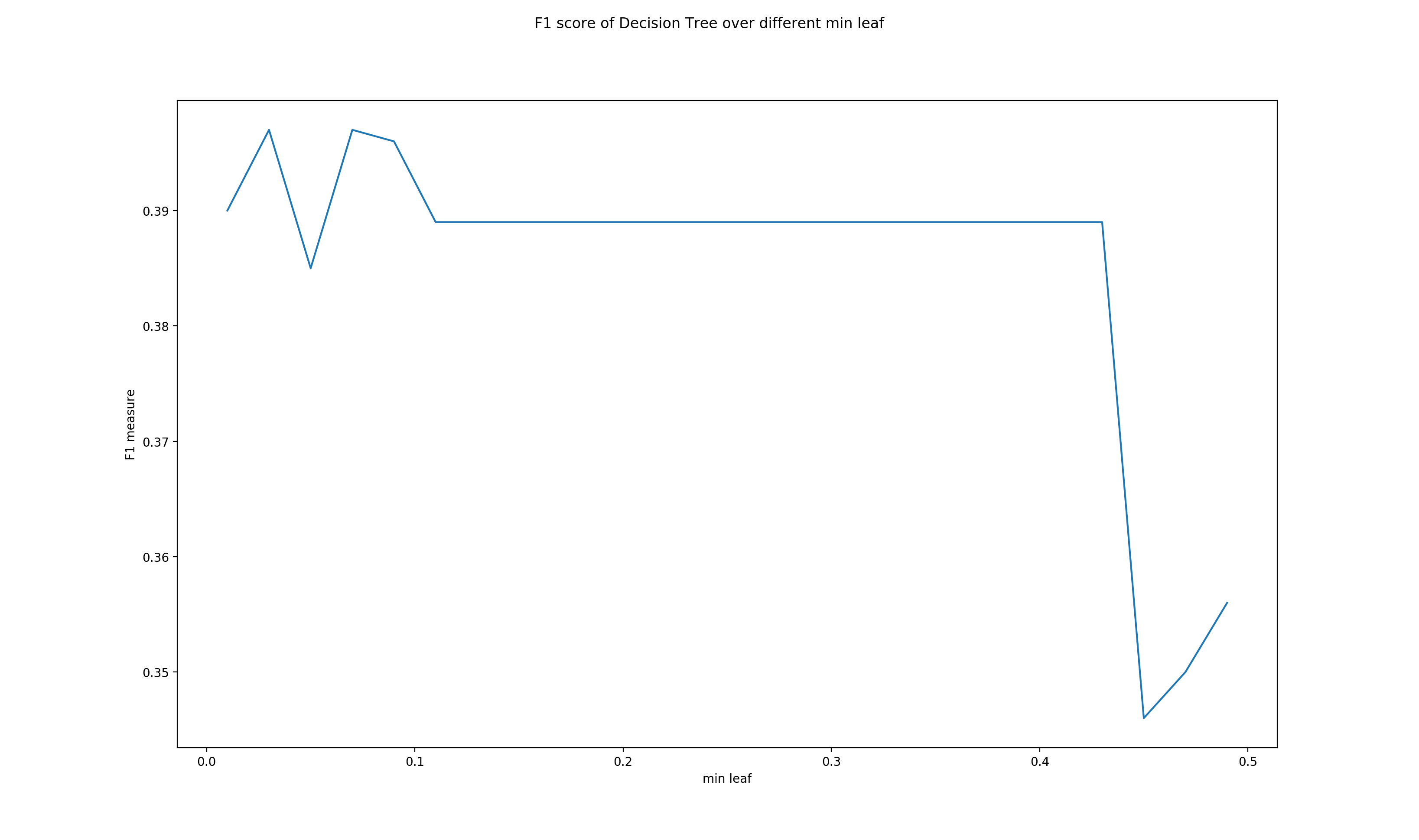
The best max depth: 11 (range :1 to 30)

**min sample split:**



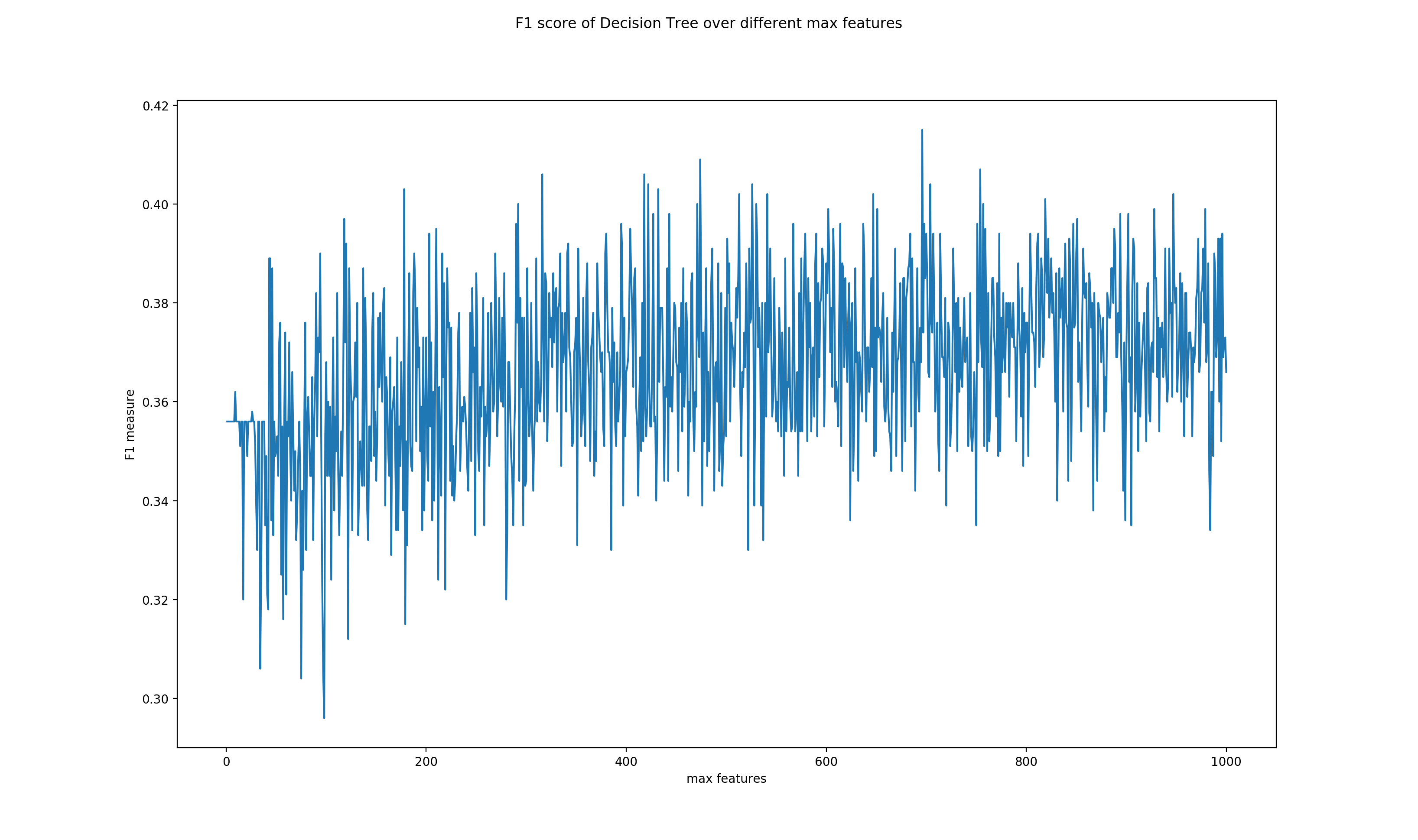
The best min split is: 0.1 (range :0 to 1)

**min sample leaf:**



The best min leaf is: 0.03 (range :0 to 0.5)

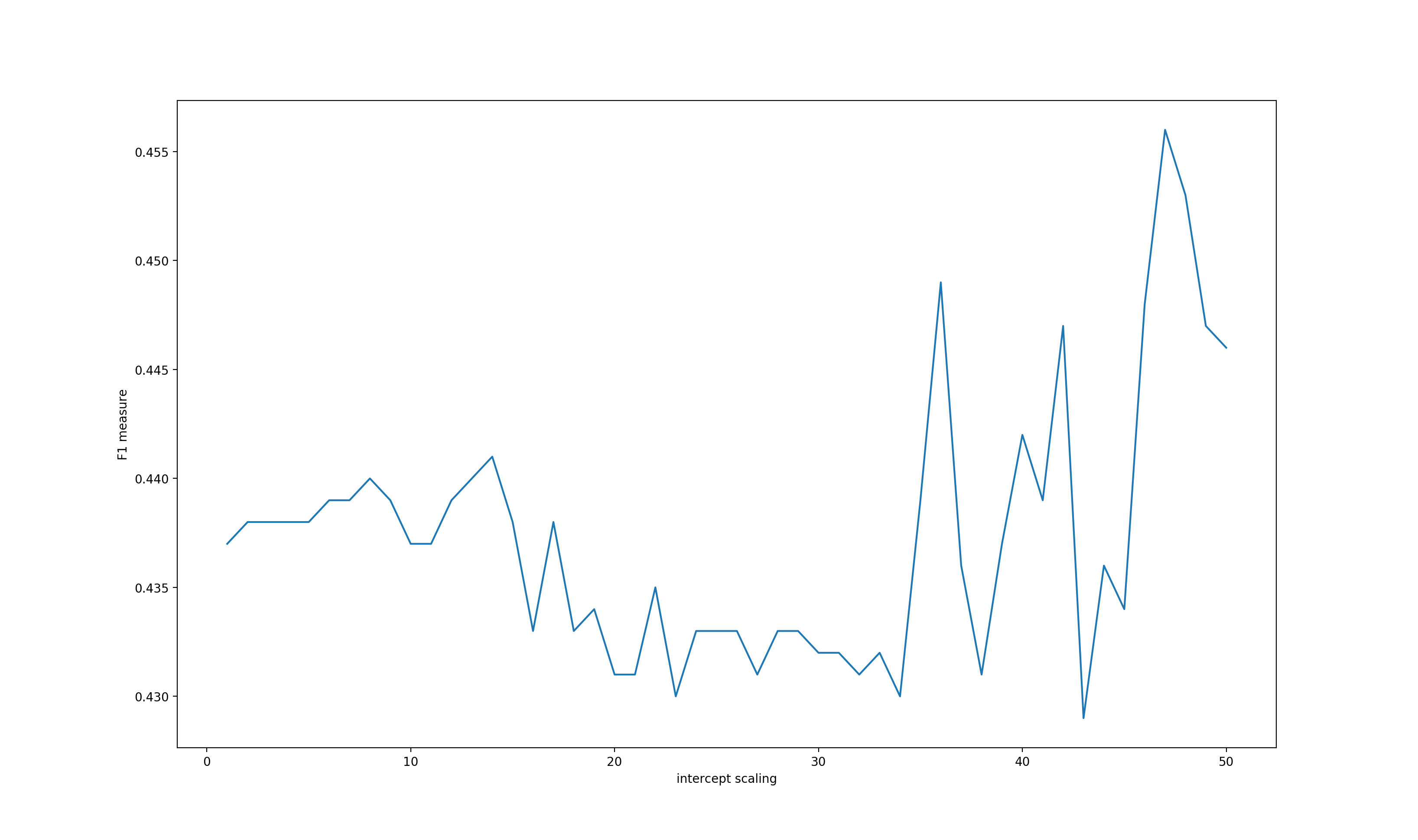
**max features:**

****

The best max features: 696 (range :1 to 1000)

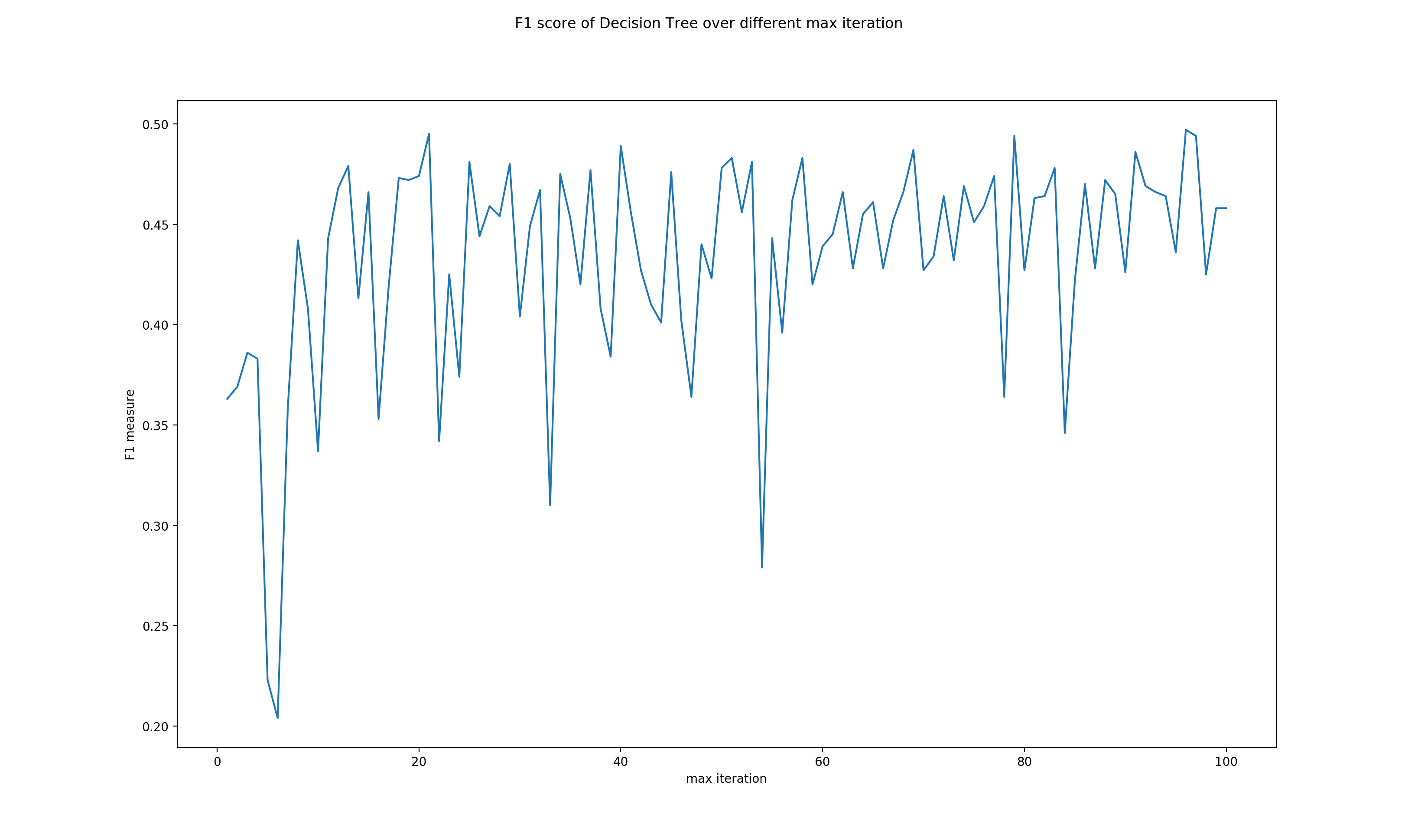
**3.** Tune linear svm

**intercept scaling:**

****

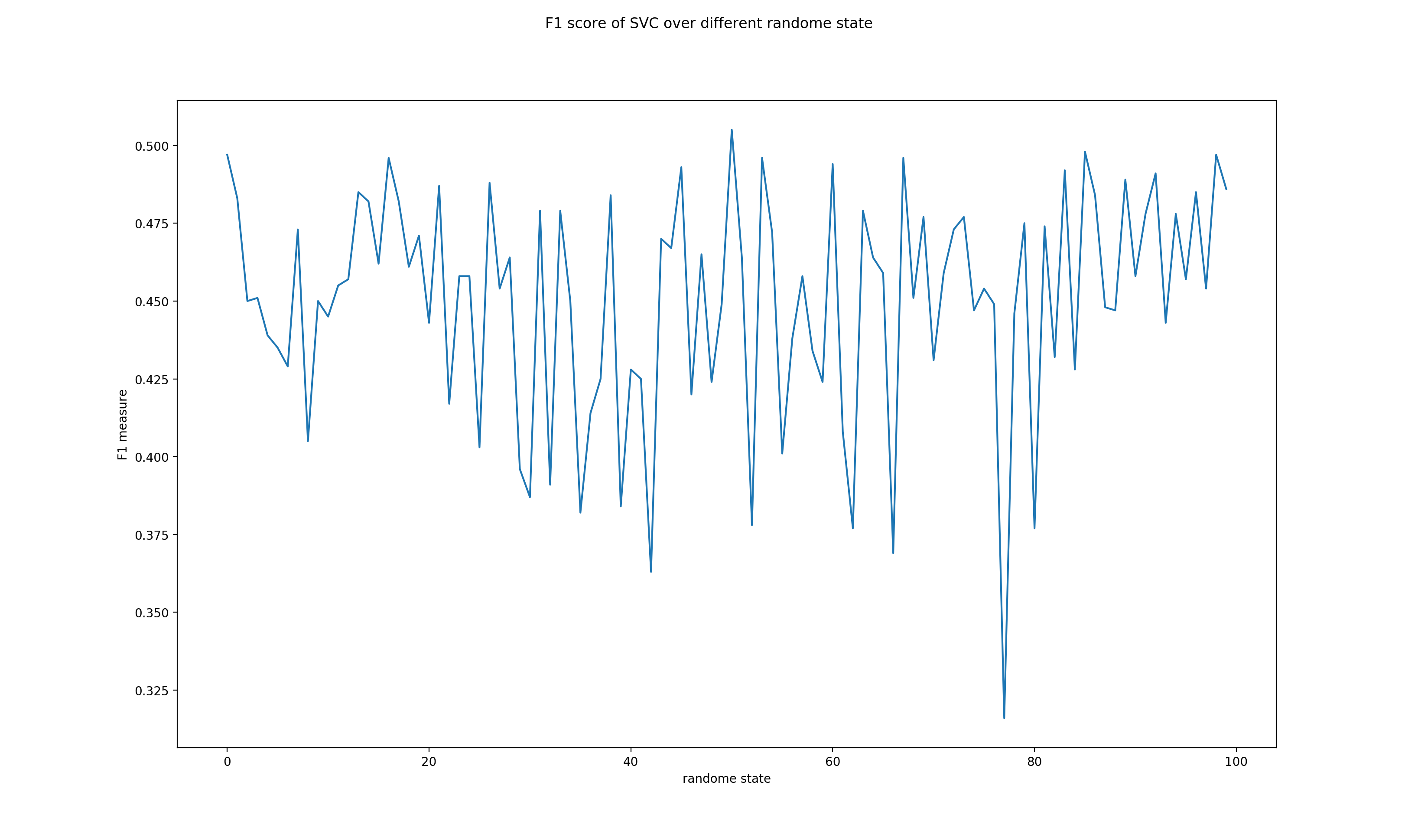
The best intercept scaling: 47 (range :1 to 50)

**max iteration:**

****

The best max iteration is: 96 (range :1 to 100)

**random state:**

****

The best random state is: 50 (range :1 to 100)

**tolerance:**

****

The best tolerance is: 1e-06 (range :1e-6 to 1e-3)

**(c)**

**Naïve Bayes:** best alpha: 0.0100 (range: 1e-10 to 0.1)

**Decision Tree:**

The best max depth: 11 (range :1 to 30)

The best min split is: 0.1 (range :0 to 1)

The best min leaf is: 0.03 (range :0 to 0.5)

The best max features: 696 (range :1 to 1000)

**Linear SVM:**

The intercept scaling: 47 (range :1 to 50)

The best max iteration is: 96 (range :1 to 100)

The best random state is: 50 (range :1 to 100)

The best tolerance is: 1e-06 (range :1e-6 to 1e-3)

These data are reported in files “Assignment3\_260540022\_2\_c\_naive\_bayes\_hyperparameter.txt”, “Assignment3\_260540022\_2\_c\_decision\_tree\_hyperparameter.txt” and “Assignment3\_260540022\_2\_c\_svm\_hyperparameter.txt”

**(d)**

Train performance:

naive bayes: 0.7305714285714285

decision tree: 0.41228571428571426

svm: 0.898857142857143

Valid performance:

naive bayes: 0.424

decision tree: 0.377

svm: 0.505

Test performance:

naive bayes: 0.444

decision tree: 0.396

svm: 0.501

These data are reported in file “Assignment\_260540022\_2\_d.txt”

**(e)**

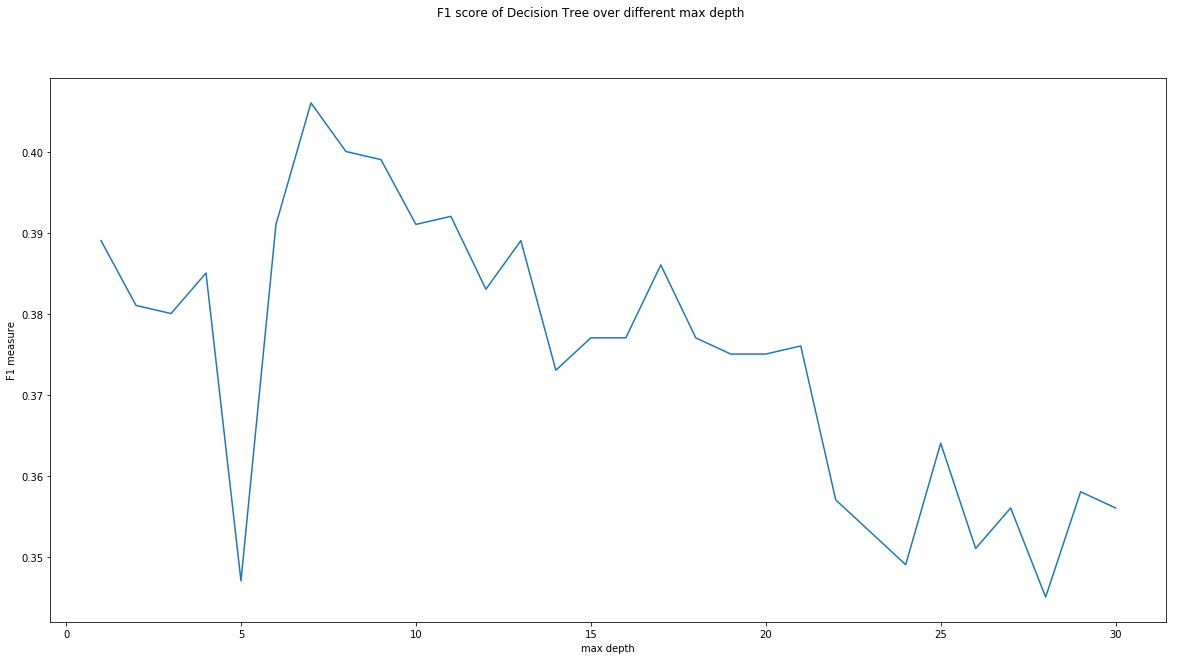
**Question 3.**

**(a)**

**1**. Gaussian Naïve Bayes will not be tuned

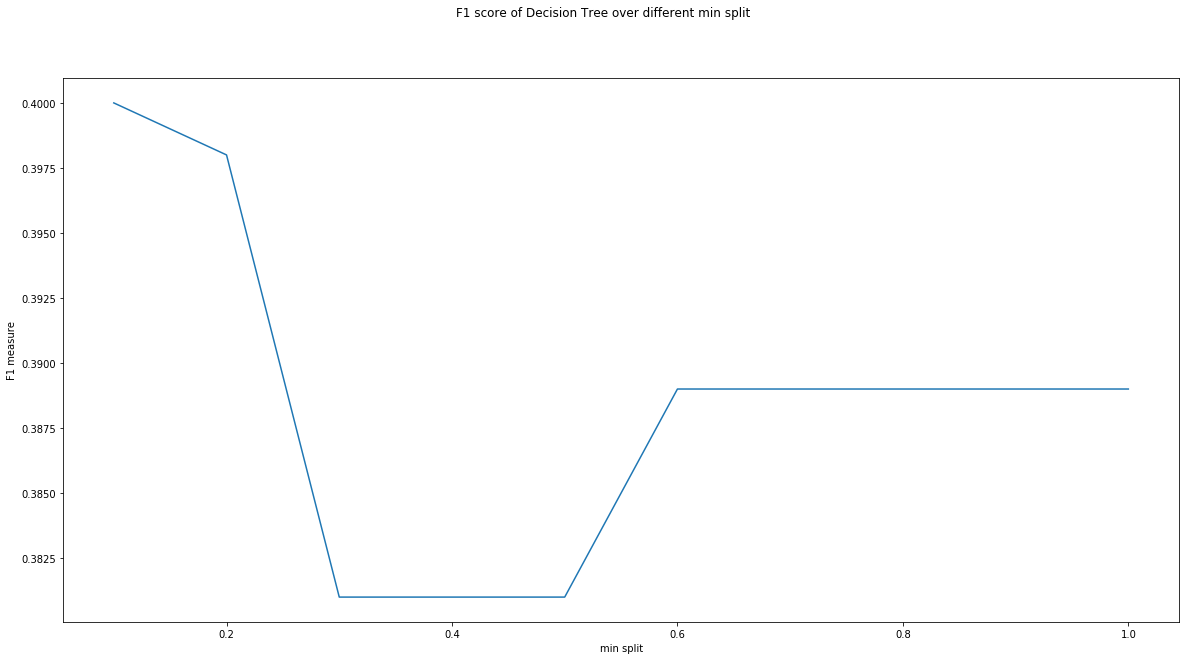
**2.** Tune Decision Tree:

**depth:**



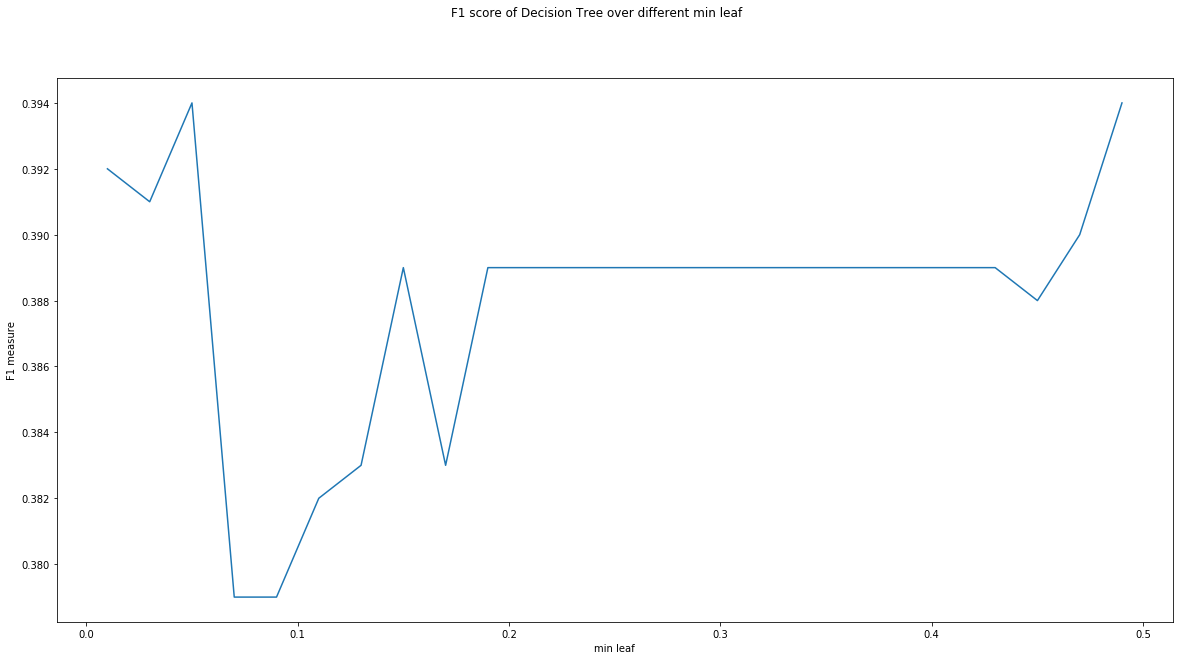
The best max depth: 7 (range :1 to 30)

**min sample split:**



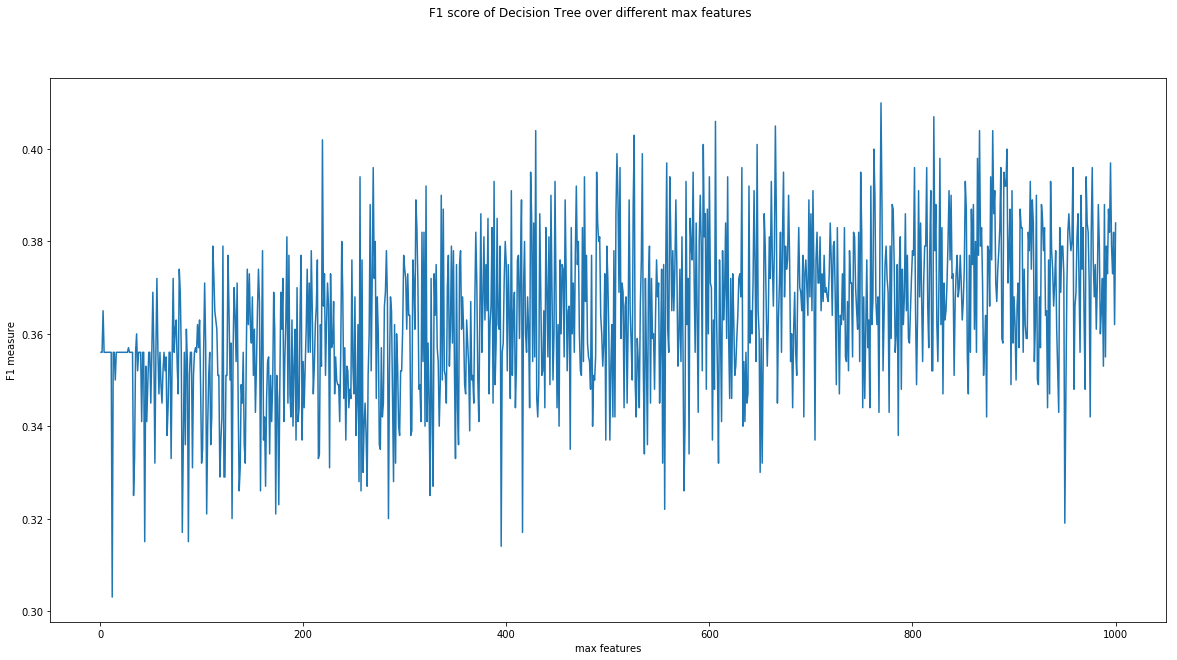
The best min split is: 0.1 (range :0 to 1)

**min sample leaf:**



The best min leaf is: 0.05 (range :0 to 0.5)

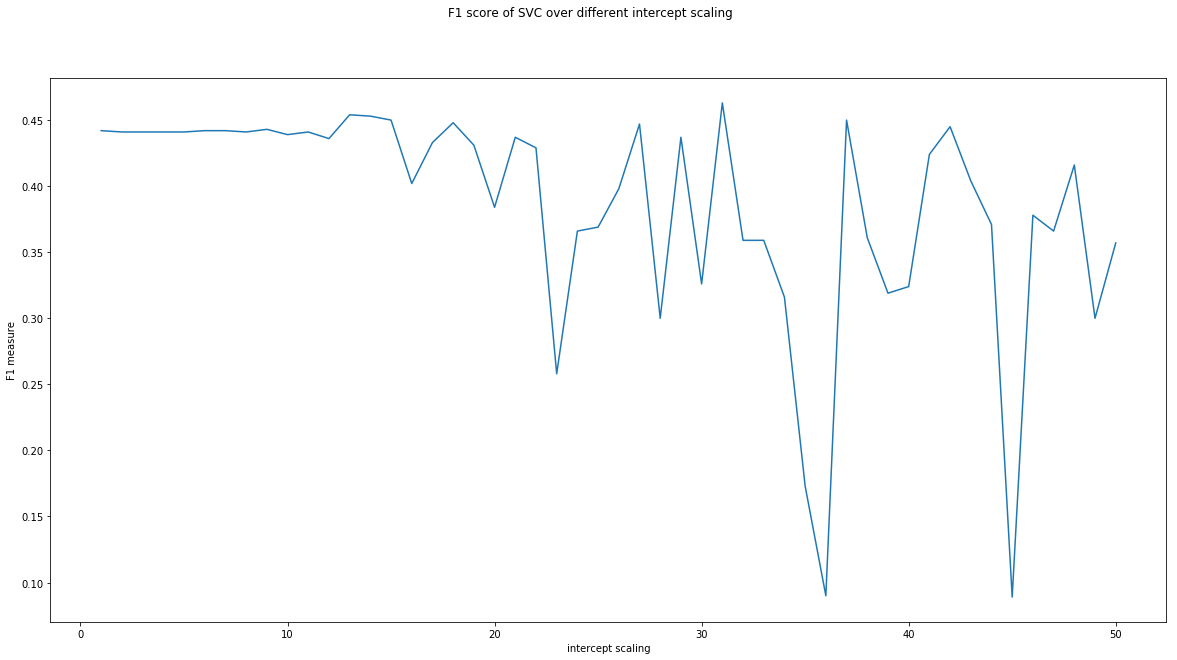
**max features:**

****

The best max features: 769 (range :1 to 1000)

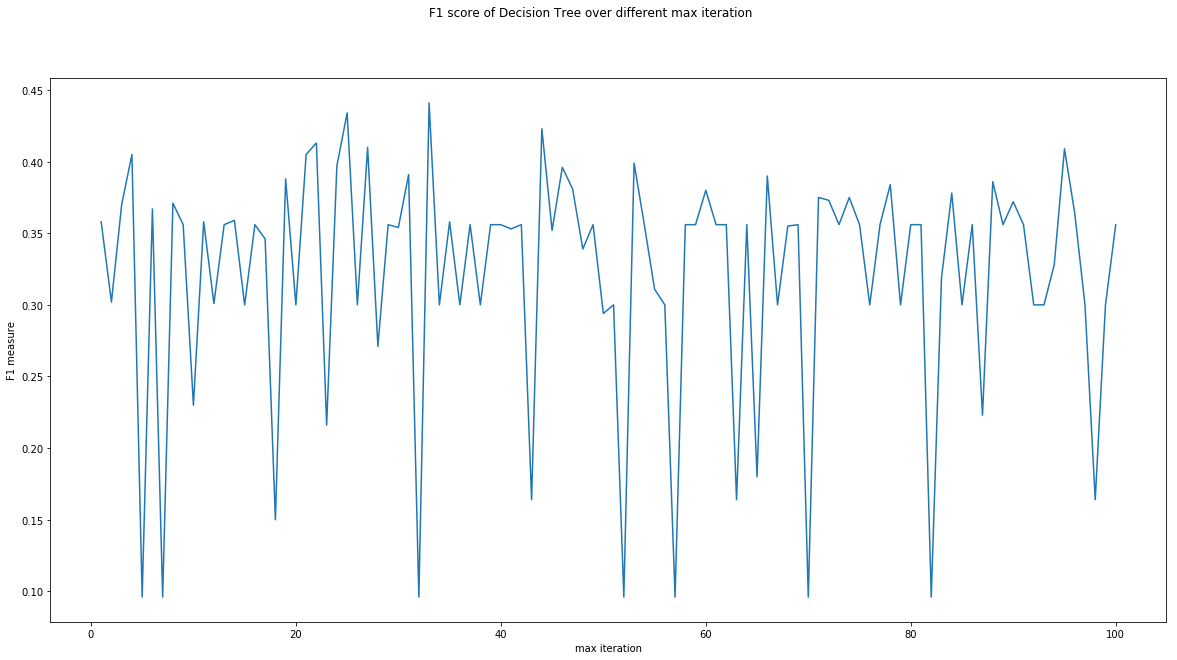
**3.** Tune linear svm

**intercept scaling:**

****

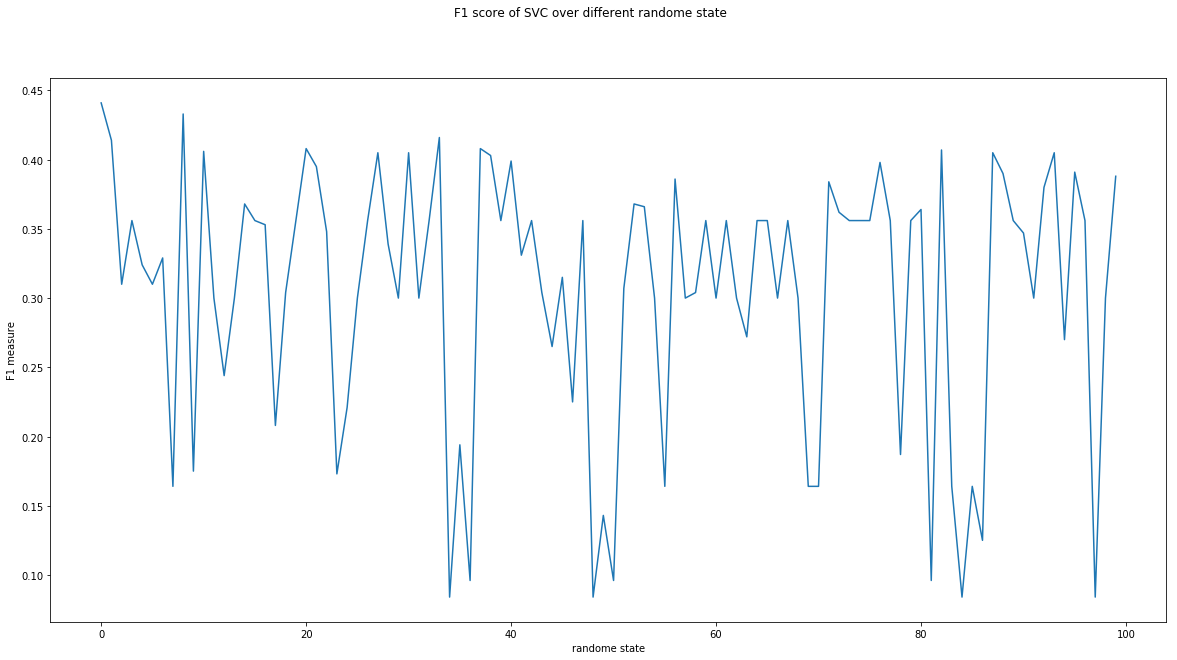
The best intercept scaling: 31 (range: 1 to 50)

**max iteration:**

****

The best max iteration: 33 (range :1 to 100)

**random state:**

****

The best max random state: 0 (range :1 to 100)

**tolerance:**

****

The best max tolerance: 1e-06 (range : 1e-6 to 1e-3)

**(c)**

Train performance:

naive bayes: 0.7937142857142857

decision tree: 0.399

svm: 0.44385714285714284

Valid performance:

naive bayes: 0.301

decision tree: 0.37

svm: 0.441

Test performance:

naive bayes: 0.32

decision tree: 0.369

svm: 0.425

These data are reported in file “Assignment\_260540022\_3\_c.txt”

**(d)**

**(e)**

**(f)**

**Question 4.**

**(a)**

Random uniform IMDB uniform: 0.50056