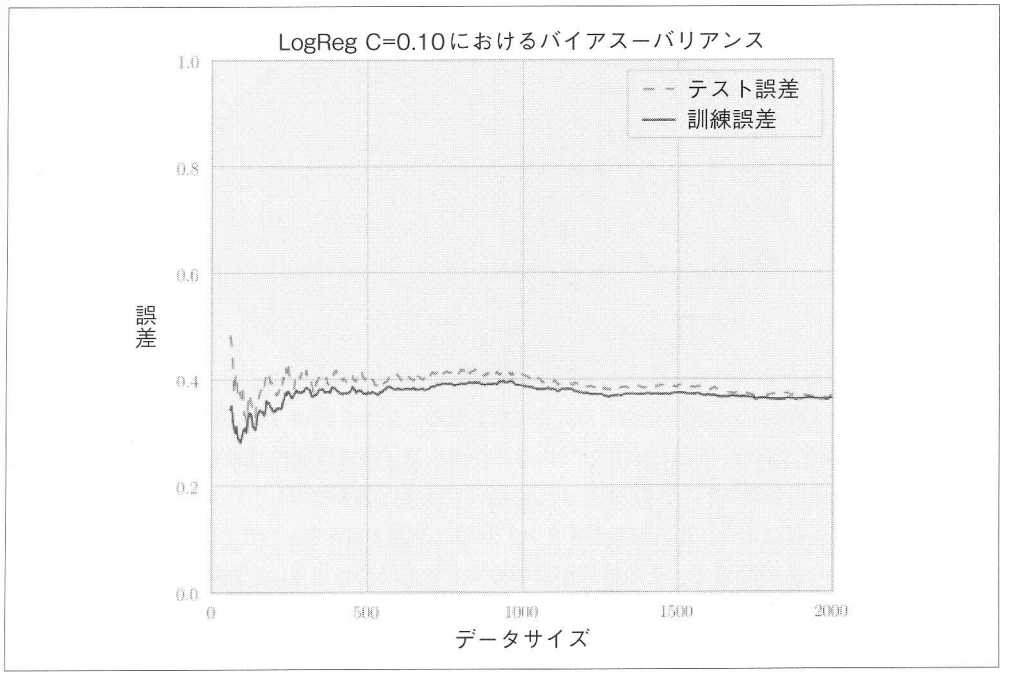

図5-10

5.7 適合率と再現率

私たちはこれまでずっと分類器の評価を行うために正解率という指標を用いてきました。ここでは今一度、私たちが達成したいことについて考えることにします。実際、私たちは回答の良し悪しを完璧に予測できる分類器を求めていません。

もし、どちらか一つのクラスだけは上手く予測できるように分類器を調整することができたら、それに応じてユーザへ適切なフィードバックを与えることができるでしょう。たとえば、「ある文書を“悪い”回答と予測した場合は、常に事実は正しい（実際に“悪い”回答である）」という能力のある分類器があったとします。そのような場合、その分類器が回答を「悪い」と予測した場合にのみ、私たちはユーザにフィードバックを与えるでしょう。これとは逆に、「ある文書を“良い”回答と予測した場合は、常に事実は正しい」という分類器の場合、初めはユーザに役立つコメントを表示しておいて、分類器が回答を「良い」と予測した段階で、そのコメントを表示しないようにすることができます。

私たちの置かれている状況をより良く理解するためには、**適合率**（precision）と**再現率**（recall）について理解する必要があります。そのために、表5-6に示される4つの分類結果（予測と事実の組み合わせ）について見ていきたいと思います。

表5-6

		分類器の予測	
		陽性	陰性
事実	陽性	True positive (TP)	False negative (FN)
	陰性	False positive (FP)	True negative (TN)

たとえば、分類器がある問題に対して「陽性（Positive）」であると予測して、事実も陽性の場合、分類器は「正しい（True）」結果を出力したことになるため、これはTrue Positiveであると言います。また逆に、分類器が「陰性（Negative）」であると予測して、事実は陽性の場合、分類器は「誤った（False）」結果を出力したことになり、False Negativeであると言います†。

私たちが求めていることは、ある回答を「良い」回答と予測するとき、もしくは「悪い」と予測するときの、どちらか一方の場合で、その結果が高い確率で正しくなるようにしたいのです。必ずしも良い・悪い**両方**の場合での予測の正しさを期待しているのではありません。ここで、たとえば、Positive用の分類能力を評価したいとします。その場合、分類器が「Positive」と予測した中で、それが正しい割合を計算すれば、それを指標として用いることができます。Postiveと予想した場合の中で実際にそれが正しい割合は**適合率**（precision）と呼ばれ、次の式で定義されます。

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$$

もし私たちの目標ができるだけ多くの良い回答（または悪い回答）を見つけることであれば、**再現率**（recall）がその場合の指標になります。再現率は、良い回答をどれだけ取りこぼしなく集めることができたか、ということを表します。

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$$

図5-11は、「全ての良い回答（事実）」と「良い回答であると分類された回答（予測）」からなる集合を表しています。

† 訳注：「陽性（Positive）」「陰性（Negative）」という言葉は、分類器の予測結果を指します。たとえば、病気を発見するための検査の場合、結果が「陽性」であるとは“疑いあり”と、「陰性」であるとは“疑いなし”と解釈できます。

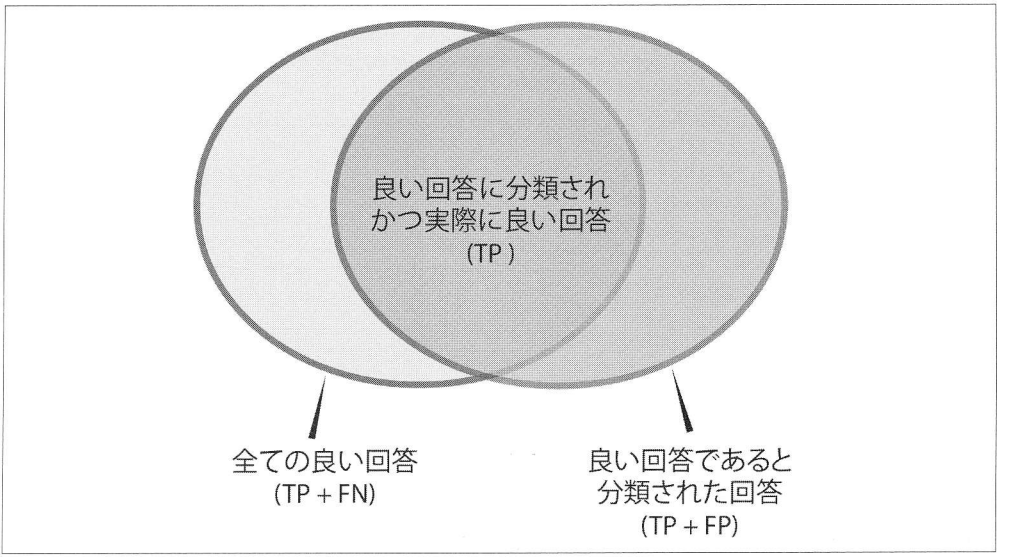

図5-11

適合率は、右の円における交差領域の割合です。再現率は、左の円における交差領域の割合です。

それでは適合率を最適にする方法を考えたいと思います。今のところ、閾値に0.5という値を用いて、回答の良し悪しを判定しています。私たちにできることは、閾値の値を変更したときのTP・FP・FNの数を数えることです。それらの値から、再現率と適合率の変移をプロットすることができます。

sklearnのmetricsモジュールには、precision_recall_curve()という関数が用意されています。この関数を使うことで、適合率と再現率の計算を行うことができます。そのためのコードの例を次に示します。

```
>>> from sklearn.metrics import precision_recall_curve
>>> precision, recall, thresholds = precision_recall_curve(y_test, clf.predict(X_
test))
```

片方のクラスを許容できる精度で分類できたとしても、もう片方のクラスを同じように許容できる精度で分類できるとは限りません。このことは次の二つのグラフから確認できます。図5-12のグラフは**適合率-再現率曲線**（Precision-Recall curve）であり、左図が悪い回答を分類する場合を、右図は良い回答の場合を示しています。