

## 演習

University of California Irvine では, “Machine Learning Repository” という機械学習のベンチマークデータセットを大量に保存・公開している. そのサイトにある “Wine Quality Data Set” をダウンロードせよ. このデータは, ポルトガルの様々な赤・白のワインの化学成分分析結果による 11 種類の特徴量と主観評価による品質指標が組み合わされたデータである.

### スクリプト (共通)

全課題に共通する, CSV の読み込み部分のスクリプトを以下に示す.

```
# の読み込み csv
dat.R = read.table("winequality-red.csv", header=T, sep=";")
dat.W = read.table("winequality-white.csv", header=T, sep=";")
```

### 相関

1.

白ワイン, 赤ワインそれぞれについて, 品質指標と最も高い相関を持つ特徴量は何か, 求めなさい.

### スクリプト

```
highest_cor <- function(dat) {
  correlation = rep(0, 11)
  for(i in 1:11) {
    correlation[i] = cor(dat$quality, dat[i])
  }
  return (names(dat)[match(max(correlation), correlation)]);
}

highest_cor(dat.R)
highest_cor(dat.W)
```

### 実行結果

上記のスクリプトを実行した結果を以下に示す.

```
> highest_cor(dat.R)
[1] "alcohol"
> highest_cor(dat.W)
[1] "alcohol"
```

また, それぞれの特徴量と品質との相関を以下に示す.

```
> correlation.R # 赤ワイン
[1] 0.12405165 -0.39055778 0.22637251 0.01373164 -0.12890656 -0.05065606
[7] -0.18510029 -0.17491923 -0.05773139 0.25139708 0.47616632
```

```
> correlation.W # 白ワイン
```

```
[1] -0.113662831 -0.194722969 -0.009209091 -0.097576829 -0.209934411  
[6]  0.008158067 -0.174737218 -0.307123313  0.099427246  0.053677877  
[11]  0.435574715
```

## 考察

赤ワイン, 白ワインともに alcohol が最も品質指標と高い相関を持つ特徴量であった。一方で, 最も小さい特徴量は赤ワインでは volatile.acidity, 白ワインでは density であった。これらは, 負の相関を持つ特徴量であり, これらの値が大きいほど品質は下がるといえる。

## 重回帰

2.

白ワイン, 赤ワインそれぞれについて, 11 種類すべての特徴量を説明変数とした 1 次式で, 品質指標を回帰しなさい。回帰の上で重要な特徴量と重要でない特徴量は何か。白ワインと赤ワインではどのような相違があるか考察しなさい。

## スクリプト

```
summary(lm(quality ~ ., data=dat.R))  
summary(lm(quality ~ ., data=dat.W))
```

## 実行結果

```
> summary(lm(quality ~ ., data=dat.R)) # 赤ワイン
```

Call:

```
lm(formula = quality ~ ., data = dat.R)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.68911	-0.36652	-0.04699	0.45202	2.02498

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.197e+01	2.119e+01	1.036	0.3002
fixed.acidity	2.499e-02	2.595e-02	0.963	0.3357
volatile.acidity	-1.084e+00	1.211e-01	-8.948	< 2e-16 ***
citric.acid	-1.826e-01	1.472e-01	-1.240	0.2150
residual.sugar	1.633e-02	1.500e-02	1.089	0.2765
chlorides	-1.874e+00	4.193e-01	-4.470	8.37e-06 ***
free.sulfur.dioxide	4.361e-03	2.171e-03	2.009	0.0447 *

total.sulfur.dioxide	-3.265e-03	7.287e-04	-4.480	8.00e-06	***
density	-1.788e+01	2.163e+01	-0.827	0.4086	
pH	-4.137e-01	1.916e-01	-2.159	0.0310	*
sulphates	9.163e-01	1.143e-01	8.014	2.13e-15	***
alcohol	2.762e-01	2.648e-02	10.429	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.648 on 1587 degrees of freedom

Multiple R-squared: 0.3606, Adjusted R-squared: 0.3561

F-statistic: 81.35 on 11 and 1587 DF, p-value: < 2.2e-16

> summary(lm(quality ~ ., data=dat.W)) # 白ワイン

Call:

lm(formula = quality ~ ., data = dat.W)

Residuals:

Min	1Q	Median	3Q	Max
-3.8348	-0.4934	-0.0379	0.4637	3.1143

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.502e+02	1.880e+01	7.987	1.71e-15	***
fixed.acidity	6.552e-02	2.087e-02	3.139	0.00171	**
volatile.acidity	-1.863e+00	1.138e-01	-16.373	< 2e-16	***
citric.acid	2.209e-02	9.577e-02	0.231	0.81759	
residual.sugar	8.148e-02	7.527e-03	10.825	< 2e-16	***
chlorides	-2.473e-01	5.465e-01	-0.452	0.65097	
free.sulfur.dioxide	3.733e-03	8.441e-04	4.422	9.99e-06	***
total.sulfur.dioxide	-2.857e-04	3.781e-04	-0.756	0.44979	
density	-1.503e+02	1.907e+01	-7.879	4.04e-15	***
pH	6.863e-01	1.054e-01	6.513	8.10e-11	***
sulphates	6.315e-01	1.004e-01	6.291	3.44e-10	***
alcohol	1.935e-01	2.422e-02	7.988	1.70e-15	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7514 on 4886 degrees of freedom

Multiple R-squared: 0.2819, Adjusted R-squared: 0.2803

F-statistic: 174.3 on 11 and 4886 DF, p-value: < 2.2e-16

## 考察

赤ワインでは, volatile.acidity, chlorides, total.sulfur.dioxide, sulphates, alcohol が特に重要な特徴量となり, 他にも free.sulfur.dioxide や pH が重要な特徴量となった. 白ワインでは, volatile.acidity, residual.sugar, free.sulfur.dioxide, density, pH, sulphates, alcohol が特に重要な特徴量となり, 他にも fixed.acidity が重要な特徴量となった. volatile.acidity や alcohol はどちらも重要だが, それ以外は異なっている. このことから, 赤ワインと白ワインの評価指標は全く異なることが分かる.

## 3.

特徴量同士の積 (クロスターム) も用いた 2 次式で回帰した場合には, 回帰の質は改善するか, 評価しなさい.

## スクリプト

```
summary(lm(quality ~ .^2, data=dat.R))
summary(lm(quality ~ .^2, data=dat.W))
```

## 実行結果

```
> summary(lm(quality ~ .^2, data=dat.R)) # 赤ワイン
```

Call:

```
lm(formula = quality ~ .^2, data = dat.R)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.29483	-0.36214	-0.06062	0.40231	1.89781

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.909e+02	5.430e+02	0.536	0.592260
fixed.acidity	-8.832e+00	1.020e+01	-0.866	0.386695
volatile.acidity	-2.226e+01	1.625e+02	-0.137	0.891041
citric.acid	-1.413e+02	1.679e+02	-0.841	0.400273
residual.sugar	1.110e+01	1.146e+01	0.969	0.332699
chlorides	-1.239e+03	6.200e+02	-1.999	0.045795 *
free.sulfur.dioxide	-8.424e+00	2.987e+00	-2.821	0.004856 **
total.sulfur.dioxide	3.403e+00	1.027e+00	3.313	0.000944 ***
density	-2.849e+02	5.444e+02	-0.523	0.600864
pH	-8.579e+01	1.482e+02	-0.579	0.562658
sulphates	3.829e+02	1.621e+02	2.363	0.018255 *
alcohol	-5.298e+00	1.242e+01	-0.427	0.669651
fixed.acidity:volatile.acidity	-1.872e-01	1.849e-01	-1.012	0.311629

fixed.acidity:citric.acid	-3.062e-01	1.822e-01	-1.680	0.093166	.
fixed.acidity:residual.sugar	1.607e-02	1.934e-02	0.831	0.406195	
fixed.acidity:chlorides	-1.804e+00	6.230e-01	-2.895	0.003839	**
fixed.acidity:free.sulfur.dioxide	-7.545e-03	3.389e-03	-2.226	0.026152	*
fixed.acidity:total.sulfur.dioxide	2.118e-03	1.278e-03	1.658	0.097604	.
fixed.acidity:density	8.316e+00	1.010e+01	0.823	0.410415	
fixed.acidity:pH	3.411e-01	1.198e-01	2.847	0.004470	**
fixed.acidity:sulphates	3.405e-01	1.720e-01	1.979	0.047944	*
fixed.acidity:alcohol	-4.357e-02	2.188e-02	-1.992	0.046593	*
volatile.acidity:citric.acid	1.091e+00	7.855e-01	1.388	0.165243	
volatile.acidity:residual.sugar	-4.365e-02	1.331e-01	-0.328	0.742949	
volatile.acidity:chlorides	1.318e+00	3.027e+00	0.436	0.663220	
volatile.acidity:free.sulfur.dioxide	-2.479e-02	1.754e-02	-1.413	0.157807	
volatile.acidity:total.sulfur.dioxide	2.151e-02	5.483e-03	3.923	9.12e-05	***
volatile.acidity:density	2.410e+01	1.651e+02	0.146	0.883933	
volatile.acidity:pH	-1.736e+00	1.245e+00	-1.394	0.163576	
volatile.acidity:sulphates	1.149e-01	9.115e-01	0.126	0.899736	
volatile.acidity:alcohol	3.470e-01	2.028e-01	1.711	0.087225	.
citric.acid:residual.sugar	-1.546e-03	1.125e-01	-0.014	0.989033	
citric.acid:chlorides	2.935e-01	2.987e+00	0.098	0.921755	
citric.acid:free.sulfur.dioxide	8.397e-03	2.031e-02	0.413	0.679389	
citric.acid:total.sulfur.dioxide	5.267e-03	6.067e-03	0.868	0.385492	
citric.acid:density	1.563e+02	1.699e+02	0.920	0.357778	
citric.acid:pH	-5.864e+00	1.434e+00	-4.088	4.58e-05	***
citric.acid:sulphates	-8.027e-02	1.146e+00	-0.070	0.944184	
citric.acid:alcohol	5.986e-01	2.126e-01	2.816	0.004927	**
residual.sugar:chlorides	-1.683e-02	5.493e-01	-0.031	0.975568	
residual.sugar:free.sulfur.dioxide	-1.320e-03	1.688e-03	-0.782	0.434226	
residual.sugar:total.sulfur.dioxide	1.315e-03	6.273e-04	2.097	0.036182	*
residual.sugar:density	-1.132e+01	1.158e+01	-0.978	0.328341	
residual.sugar:pH	9.004e-02	2.142e-01	0.420	0.674224	
residual.sugar:sulphates	1.403e-01	1.582e-01	0.887	0.375487	
residual.sugar:alcohol	-3.167e-02	2.123e-02	-1.492	0.135974	
chlorides:free.sulfur.dioxide	-1.215e-01	7.345e-02	-1.654	0.098354	.
chlorides:total.sulfur.dioxide	-5.564e-04	2.868e-02	-0.019	0.984524	
chlorides:density	1.264e+03	6.267e+02	2.018	0.043803	*
chlorides:pH	-3.724e+00	5.635e+00	-0.661	0.508740	
chlorides:sulphates	-1.935e+00	2.599e+00	-0.744	0.456805	
chlorides:alcohol	6.764e-01	8.565e-01	0.790	0.429795	
free.sulfur.dioxide:total.sulfur.dioxide	-1.097e-04	5.383e-05	-2.037	0.041788	*
free.sulfur.dioxide:density	8.582e+00	3.044e+00	2.820	0.004867	**

free.sulfur.dioxide:pH	-3.427e-02	2.339e-02	-1.465	0.143129	
free.sulfur.dioxide:sulphates	-4.412e-02	1.677e-02	-2.632	0.008580	**
free.sulfur.dioxide:alcohol	1.069e-02	3.938e-03	2.715	0.006693	**
total.sulfur.dioxide:density	-3.459e+00	1.050e+00	-3.293	0.001013	**
total.sulfur.dioxide:pH	1.524e-02	8.892e-03	1.714	0.086651	.
total.sulfur.dioxide:sulphates	5.908e-03	5.378e-03	1.098	0.272184	
total.sulfur.dioxide:alcohol	-4.434e-03	1.416e-03	-3.130	0.001780	**
density:pH	8.413e+01	1.482e+02	0.568	0.570386	
density:sulphates	-3.933e+02	1.641e+02	-2.397	0.016639	*
density:alcohol	5.919e+00	1.273e+01	0.465	0.642086	
pH:sulphates	2.625e+00	1.195e+00	2.196	0.028271	*
pH:alcohol	-5.888e-02	2.018e-01	-0.292	0.770495	
sulphates:alcohol	-8.614e-02	1.998e-01	-0.431	0.666456	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.625 on 1532 degrees of freedom

Multiple R-squared: 0.4258, Adjusted R-squared: 0.401

F-statistic: 17.21 on 66 and 1532 DF, p-value: < 2.2e-16

> summary(lm(quality ~ .^2, data=dat.W)) # 白ワイン

Call:

lm(formula = quality ~ .^2, data = dat.W)

Residuals:

Min	1Q	Median	3Q	Max
-3.3803	-0.4855	-0.0096	0.4340	3.1062

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-1.600e+02	3.506e+02	-0.456	0.64821	
fixed.acidity	2.326e+01	1.520e+01	1.530	0.12612	
volatile.acidity	-4.695e+02	1.464e+02	-3.207	0.00135	**
citric.acid	-1.826e+02	1.851e+02	-0.986	0.32397	
residual.sugar	1.203e+00	9.464e-01	1.271	0.20383	
chlorides	-2.999e+03	1.237e+03	-2.424	0.01539	*
free.sulfur.dioxide	-3.088e+00	1.613e+00	-1.914	0.05562	.
total.sulfur.dioxide	-7.726e-01	6.075e-01	-1.272	0.20346	
density	1.686e+02	3.513e+02	0.480	0.63118	
pH	1.461e+02	9.405e+01	1.553	0.12042	

sulphates	-1.354e+01	1.751e+02	-0.077	0.93839	
alcohol	2.254e+01	7.608e+00	2.963	0.00306	**
fixed.acidity:volatile.acidity	-4.883e-01	2.008e-01	-2.432	0.01505	*
fixed.acidity:citric.acid	-3.459e-01	1.859e-01	-1.861	0.06281	.
fixed.acidity:residual.sugar	1.134e-02	6.176e-03	1.837	0.06632	.
fixed.acidity:chlorides	-1.938e+00	1.340e+00	-1.446	0.14817	
fixed.acidity:free.sulfur.dioxide	7.355e-04	1.681e-03	0.438	0.66168	
fixed.acidity:total.sulfur.dioxide	-6.188e-04	6.917e-04	-0.895	0.37108	
fixed.acidity:density	-2.366e+01	1.515e+01	-1.562	0.11844	
fixed.acidity:pH	2.749e-01	8.855e-02	3.104	0.00192	**
fixed.acidity:sulphates	2.022e-01	1.822e-01	1.110	0.26704	
fixed.acidity:alcohol	-2.381e-02	2.623e-02	-0.908	0.36401	
volatile.acidity:citric.acid	1.127e+00	8.318e-01	1.354	0.17564	
volatile.acidity:residual.sugar	-1.612e-01	6.049e-02	-2.666	0.00771	**
volatile.acidity:chlorides	2.064e-01	6.008e+00	0.034	0.97259	
volatile.acidity:free.sulfur.dioxide	9.695e-03	8.598e-03	1.128	0.25956	
volatile.acidity:total.sulfur.dioxide	1.021e-03	3.505e-03	0.291	0.77093	
volatile.acidity:density	4.604e+02	1.489e+02	3.093	0.00200	**
volatile.acidity:pH	7.809e-01	1.087e+00	0.718	0.47267	
volatile.acidity:sulphates	-7.225e-01	1.032e+00	-0.700	0.48404	
volatile.acidity:alcohol	1.102e+00	1.901e-01	5.798	7.15e-09	***
citric.acid:residual.sugar	-9.546e-02	6.895e-02	-1.384	0.16629	
citric.acid:chlorides	-8.101e-01	3.833e+00	-0.211	0.83264	
citric.acid:free.sulfur.dioxide	7.037e-03	7.113e-03	0.989	0.32256	
citric.acid:total.sulfur.dioxide	-2.792e-03	2.902e-03	-0.962	0.33606	
citric.acid:density	1.817e+02	1.874e+02	0.970	0.33213	
citric.acid:pH	8.733e-01	9.587e-01	0.911	0.36236	
citric.acid:sulphates	-7.540e-01	9.453e-01	-0.798	0.42515	
citric.acid:alcohol	2.195e-01	2.358e-01	0.931	0.35211	
residual.sugar:chlorides	-1.486e+00	4.719e-01	-3.149	0.00165	**
residual.sugar:free.sulfur.dioxide	-1.603e-03	6.332e-04	-2.532	0.01137	*
residual.sugar:total.sulfur.dioxide	-1.807e-04	2.494e-04	-0.725	0.46866	
residual.sugar:density	-9.535e-01	9.186e-01	-1.038	0.29931	
residual.sugar:pH	-1.987e-02	3.491e-02	-0.569	0.56925	
residual.sugar:sulphates	-4.399e-02	6.799e-02	-0.647	0.51773	
residual.sugar:alcohol	5.090e-03	4.639e-03	1.097	0.27259	
chlorides:free.sulfur.dioxide	4.086e-02	4.036e-02	1.012	0.31138	
chlorides:total.sulfur.dioxide	-2.274e-02	2.333e-02	-0.975	0.32968	
chlorides:density	3.086e+03	1.250e+03	2.469	0.01359	*
chlorides:pH	-1.893e+01	6.265e+00	-3.021	0.00253	**
chlorides:sulphates	-1.404e+01	6.347e+00	-2.212	0.02698	*

chlorides:alcohol	1.836e+00	1.641e+00	1.119	0.26330
free.sulfur.dioxide:total.sulfur.dioxide	-9.951e-05	9.267e-06	-10.738	< 2e-16 ***
free.sulfur.dioxide:density	3.039e+00	1.634e+00	1.860	0.06296 .
free.sulfur.dioxide:pH	3.414e-03	8.438e-03	0.405	0.68580
free.sulfur.dioxide:sulphates	2.955e-02	7.314e-03	4.040	5.43e-05 ***
free.sulfur.dioxide:alcohol	5.765e-03	2.174e-03	2.653	0.00801 **
total.sulfur.dioxide:density	8.022e-01	6.171e-01	1.300	0.19366
total.sulfur.dioxide:pH	-5.218e-03	3.856e-03	-1.353	0.17605
total.sulfur.dioxide:sulphates	-1.667e-02	3.605e-03	-4.624	3.86e-06 ***
total.sulfur.dioxide:alcohol	9.733e-04	8.117e-04	1.199	0.23056
density:pH	-1.478e+02	9.401e+01	-1.572	0.11598
density:sulphates	1.182e+01	1.773e+02	0.067	0.94684
density:alcohol	-2.318e+01	7.802e+00	-2.971	0.00299 **
pH:sulphates	1.641e+00	9.192e-01	1.785	0.07425 .
pH:alcohol	4.702e-03	1.471e-01	0.032	0.97450
sulphates:alcohol	-1.583e-01	2.195e-01	-0.721	0.47087

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7162 on 4831 degrees of freedom

Multiple R-squared: 0.3548, Adjusted R-squared: 0.346

F-statistic: 40.26 on 66 and 4831 DF, p-value: < 2.2e-16

## 考察

Multiple R-squared は決定係数であり，説明変数が目的変数をどのくらい説明できているかを表す値である．Adjusted R-squared は，自由度修正済決定係数である．これらは，1 に近づくほど良い値である．ここでは，Adjusted R-squared に着目する．

赤ワインでは，1 次式で回帰した場合，Adjusted R-squared の値は 0.3561 となっていたのに対し，2 次式で回帰した場合は，Adjusted R-squared の値は 0.401 となっている．また，白ワインでは，1 次式で回帰した場合，Adjusted R-squared の値は 0.2803 となっていたのに対し，2 次式で回帰した場合は，Adjusted R-squared の値は 0.346 となっている．したがって，赤ワイン，白ワインのいずれにおいても回帰の質は改善されているといえる．

赤ワインについては，total.sulfur.dioxide, volatile.acidity : total.sulfur.dioxide, citric.acid : pH が特に重要な特徴量となっていた．また，白ワインでは，volatile.acidity : alcohol, free.sulfur.dioxide : total.sulfur.dioxide, free.sulfur.dioxide : sulphates, total.sulfur.dioxide : sulphates の特徴量が特に重要となっていた．赤ワインについては 1 次式の際に重視されていた total.sulfur.dioxide や volatile.acidity が重視されていることが分かる．また，白ワインについても 1 次式の際に重視されていた volatile.acidity や free.sulfur.dioxide, sulphates が重視されていることが分かる．一方で，total.sulfur.dioxide など，組み合わせることによって重要度が変化した特徴量があることも確認できる．



## モデル選択

4.

特徴量の1次式を用いた回帰において、AICを用いた特徴量選択を行いたい。Rに準備されているstep関数について調べ、それを用いて最適な回帰モデルを導きなさい。

### スクリプト

```
res <- step(lm(quality ~ ., data=dat.R))
summary(res)
round(coefficients(res), 3)
res <- step(lm(quality ~ ., data=dat.W))
summary(res)
round(coefficients(res), 3)
```

### 実行結果

```
> res <- step(lm(quality ~ ., data=dat.R)) # 赤ワイン
```

Start: AIC=-1375.49

```
quality ~ fixed.acidity + volatile.acidity + citric.acid + residual.sugar +
          chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
          density + pH + sulphates + alcohol
```

	Df	Sum of Sq	RSS	AIC
- density	1	0.287	666.70	-1376.8
- fixed.acidity	1	0.389	666.80	-1376.5
- residual.sugar	1	0.498	666.91	-1376.3
- citric.acid	1	0.646	667.06	-1375.9
<none>			666.41	-1375.5
- free.sulfur.dioxide	1	1.694	668.10	-1373.4
- pH	1	1.957	668.37	-1372.8
- chlorides	1	8.391	674.80	-1357.5
- total.sulfur.dioxide	1	8.427	674.84	-1357.4
- sulphates	1	26.971	693.38	-1314.0
- volatile.acidity	1	33.620	700.03	-1298.8
- alcohol	1	45.672	712.08	-1271.5

Step: AIC=-1376.8

```
quality ~ fixed.acidity + volatile.acidity + citric.acid + residual.sugar +
          chlorides + free.sulfur.dioxide + total.sulfur.dioxide +
          pH + sulphates + alcohol
```

	Df	Sum of Sq	RSS	AIC
--	----	-----------	-----	-----

- fixed.acidity	1	0.108	666.81	-1378.5
- residual.sugar	1	0.231	666.93	-1378.2
- citric.acid	1	0.654	667.35	-1377.2
<none>			666.70	-1376.8
- free.sulfur.dioxide	1	1.829	668.53	-1374.4
- pH	1	4.325	671.02	-1368.5
- total.sulfur.dioxide	1	8.728	675.43	-1358.0
- chlorides	1	8.761	675.46	-1357.9
- sulphates	1	27.287	693.98	-1314.7
- volatile.acidity	1	35.000	701.70	-1297.0
- alcohol	1	119.669	786.37	-1114.8

Step: AIC=-1378.54

quality ~ volatile.acidity + citric.acid + residual.sugar + chlorides +  
 free.sulfur.dioxide + total.sulfur.dioxide + pH + sulphates +  
 alcohol

	Df	Sum of Sq	RSS	AIC
- residual.sugar	1	0.257	667.06	-1379.9
- citric.acid	1	0.565	667.37	-1379.2
<none>			666.81	-1378.5
- free.sulfur.dioxide	1	1.901	668.71	-1376.0
- pH	1	7.065	673.87	-1363.7
- chlorides	1	9.940	676.75	-1356.9
- total.sulfur.dioxide	1	10.031	676.84	-1356.7
- sulphates	1	27.673	694.48	-1315.5
- volatile.acidity	1	36.234	703.04	-1295.9
- alcohol	1	120.633	787.44	-1114.7

Step: AIC=-1379.93

quality ~ volatile.acidity + citric.acid + chlorides + free.sulfur.dioxide +  
 total.sulfur.dioxide + pH + sulphates + alcohol

	Df	Sum of Sq	RSS	AIC
- citric.acid	1	0.475	667.54	-1380.8
<none>			667.06	-1379.9
- free.sulfur.dioxide	1	2.064	669.13	-1377.0
- pH	1	7.138	674.20	-1364.9
- total.sulfur.dioxide	1	9.828	676.89	-1358.5
- chlorides	1	9.832	676.89	-1358.5
- sulphates	1	27.446	694.51	-1317.5

```
- volatile.acidity      1      35.977 703.04 -1297.9
- alcohol                1      122.667 789.73 -1112.0
```

Step: AIC=-1380.79

```
quality ~ volatile.acidity + chlorides + free.sulfur.dioxide +
      total.sulfur.dioxide + pH + sulphates + alcohol
```

	Df	Sum of Sq	RSS	AIC
<none>			667.54	-1380.8
- free.sulfur.dioxide	1	2.394	669.93	-1377.1
- pH	1	7.073	674.61	-1365.9
- total.sulfur.dioxide	1	10.787	678.32	-1357.2
- chlorides	1	10.809	678.35	-1357.1
- sulphates	1	27.060	694.60	-1319.2
- volatile.acidity	1	42.318	709.85	-1284.5
- alcohol	1	124.483	792.02	-1109.4

```
> summary(res)
```

Call:

```
lm(formula = quality ~ volatile.acidity + chlorides + free.sulfur.dioxide +
      total.sulfur.dioxide + pH + sulphates + alcohol, data = dat.R)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.68918	-0.36757	-0.04653	0.46081	2.02954

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.4300987	0.4029168	10.995	< 2e-16 ***
volatile.acidity	-1.0127527	0.1008429	-10.043	< 2e-16 ***
chlorides	-2.0178138	0.3975417	-5.076	4.31e-07 ***
free.sulfur.dioxide	0.0050774	0.0021255	2.389	0.017 *
total.sulfur.dioxide	-0.0034822	0.0006868	-5.070	4.43e-07 ***
pH	-0.4826614	0.1175581	-4.106	4.23e-05 ***
sulphates	0.8826651	0.1099084	8.031	1.86e-15 ***
alcohol	0.2893028	0.0167958	17.225	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6477 on 1591 degrees of freedom

Multiple R-squared: 0.3595, Adjusted R-squared: 0.3567

F-statistic: 127.6 on 7 and 1591 DF, p-value: < 2.2e-16

```
> round(coefficients(res), 3)
```

	(Intercept)	volatile.acidity	chlorides	free.sulfur.dioxide
	4.430	-1.013	-2.018	0.005
total.sulfur.dioxide		pH	sulphates	alcohol
	-0.003	-0.483	0.883	0.289

```
> res <- step(lm(quality ~ ., data=dat.W)) # 白ワイン
```

Start: AIC=-2788.44

```
quality ~ fixed.acidity + volatile.acidity + citric.acid + residual.sugar +  
          chlorides + free.sulfur.dioxide + total.sulfur.dioxide +  
          density + pH + sulphates + alcohol
```

	Df	Sum of Sq	RSS	AIC
- citric.acid	1	0.030	2758.4	-2790.4
- chlorides	1	0.116	2758.4	-2790.2
- total.sulfur.dioxide	1	0.323	2758.7	-2789.9
<none>			2758.3	-2788.4
- fixed.acidity	1	5.562	2763.9	-2780.6
- free.sulfur.dioxide	1	11.039	2769.4	-2770.9
- sulphates	1	22.339	2780.7	-2750.9
- pH	1	23.948	2782.3	-2748.1
- density	1	35.044	2793.4	-2728.6
- alcohol	1	36.020	2794.3	-2726.9
- residual.sugar	1	66.152	2824.5	-2674.4
- volatile.acidity	1	151.345	2909.7	-2528.8

Step: AIC=-2790.39

```
quality ~ fixed.acidity + volatile.acidity + residual.sugar +  
          chlorides + free.sulfur.dioxide + total.sulfur.dioxide +  
          density + pH + sulphates + alcohol
```

	Df	Sum of Sq	RSS	AIC
- chlorides	1	0.105	2758.5	-2792.2
- total.sulfur.dioxide	1	0.315	2758.7	-2791.8
<none>			2758.4	-2790.4
- fixed.acidity	1	5.749	2764.1	-2782.2
- free.sulfur.dioxide	1	11.096	2769.4	-2772.7
- sulphates	1	22.444	2780.8	-2752.7
- pH	1	23.971	2782.3	-2750.0
- density	1	35.066	2793.4	-2730.5
- alcohol	1	36.540	2794.9	-2727.9

```
- residual.sugar      1    66.160 2824.5 -2676.3
- volatile.acidity    1   156.805 2915.2 -2521.6
```

Step: AIC=-2792.2

```
quality ~ fixed.acidity + volatile.acidity + residual.sugar +
      free.sulfur.dioxide + total.sulfur.dioxide + density + pH +
      sulphates + alcohol
```

	Df	Sum of Sq	RSS	AIC
- total.sulfur.dioxide	1	0.320	2758.8	-2793.6
<none>			2758.5	-2792.2
- fixed.acidity	1	6.157	2764.6	-2783.3
- free.sulfur.dioxide	1	11.036	2769.5	-2774.7
- sulphates	1	22.570	2781.0	-2754.3
- pH	1	25.297	2783.8	-2749.5
- alcohol	1	36.536	2795.0	-2729.8
- density	1	36.823	2795.3	-2729.2
- residual.sugar	1	70.134	2828.6	-2671.2
- volatile.acidity	1	158.543	2917.0	-2520.5

Step: AIC=-2793.63

```
quality ~ fixed.acidity + volatile.acidity + residual.sugar +
      free.sulfur.dioxide + density + pH + sulphates + alcohol
```

	Df	Sum of Sq	RSS	AIC
<none>			2758.8	-2793.6
- fixed.acidity	1	6.270	2765.1	-2784.5
- free.sulfur.dioxide	1	13.826	2772.6	-2771.2
- sulphates	1	22.303	2781.1	-2756.2
- pH	1	25.460	2784.2	-2750.6
- alcohol	1	36.300	2795.1	-2731.6
- density	1	39.920	2798.7	-2725.3
- residual.sugar	1	72.942	2831.7	-2667.8
- volatile.acidity	1	167.753	2926.5	-2506.5

```
> summary(res)
```

Call:

```
lm(formula = quality ~ fixed.acidity + volatile.acidity + residual.sugar +
      free.sulfur.dioxide + density + pH + sulphates + alcohol,
      data = dat.W)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.8246	-0.4938	-0.0396	0.4660	3.1208

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.541e+02	1.810e+01	8.514	< 2e-16 ***
fixed.acidity	6.810e-02	2.043e-02	3.333	0.000864 ***
volatile.acidity	-1.888e+00	1.095e-01	-17.242	< 2e-16 ***
residual.sugar	8.285e-02	7.287e-03	11.370	< 2e-16 ***
free.sulfur.dioxide	3.349e-03	6.766e-04	4.950	7.67e-07 ***
density	-1.543e+02	1.834e+01	-8.411	< 2e-16 ***
pH	6.942e-01	1.034e-01	6.717	2.07e-11 ***
sulphates	6.285e-01	9.997e-02	6.287	3.52e-10 ***
alcohol	1.932e-01	2.408e-02	8.021	1.31e-15 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7512 on 4889 degrees of freedom

Multiple R-squared: 0.2818, Adjusted R-squared: 0.2806

F-statistic: 239.7 on 8 and 4889 DF, p-value: < 2.2e-16

> round(coefficients(res), 3)

(Intercept)	fixed.acidity	volatile.acidity	residual.sugar
154.106	0.068	-1.888	0.083
free.sulfur.dioxide	density	pH	sulphates
0.003	-154.291	0.694	0.629
alcohol			
0.193			

考察

step 関数より導かれた, quality の回帰モデルは以下の式となる.

赤ワイン

$$\text{quality}_R = 4.430 - 1.013\text{volatile.acidity} - 2.018\text{chlorides} + 0.005\text{free.sulfur.dioxide} \\ - 0.003\text{total.sulfur.dioxide} - 0.483\text{pH} + 0.883\text{sulphates} + 0.289\text{alcohol}$$

白ワイン

$$\text{quality}_W = 154.106 + 0.068\text{fixed.acidity} - 1.888\text{volatile.acidity} + 0.083\text{residual.sugar} \\ + 0.003\text{free.sulfur.dioxide} - 154.291\text{density} + 0.694\text{pH} + 0.629\text{sulphates} \\ + 0.193\text{alcohol}$$

それぞれの式に含まれる特徴量は, 2 で求めた重要な特徴量と一致している.