

支持向量机及其应用

```
#install.packages('e1071')  
library(e1071)  
library(plyr)  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':  
##  
##      arrange, count, desc, failwith, id, mutate, rename, summarise,  
##      summarize
```

```
## The following objects are masked from 'package:stats':  
##  
##      filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##      intersect, setdiff, setequal, union
```

```
library(chemCal)  
"done"
```

```
## [1] "done"
```

1. 实验一

使用支持向量机，根据收入及年龄分类，判断用户是否具有购买欲望

a) 将 Social_Network_Ads 数据导入编程语言

```
ads <- read.csv(file = '../data/Social_Network_Ads.csv')
ads = ads %>% mutate(Gender = as.factor(Gender ))
ads
```

##	User. ID	Gender	Age	EstimatedSalary	Purchased
## 1	15624510	Male	19	19000	0
## 2	15810944	Male	35	20000	0
## 3	15668575	Female	26	43000	0
## 4	15603246	Female	27	57000	0
## 5	15804002	Male	19	76000	0
## 6	15728773	Male	27	58000	0
## 7	15598044	Female	27	84000	0
## 8	15694829	Female	32	150000	1
## 9	15600575	Male	25	33000	0
## 10	15727311	Female	35	65000	0
## 11	15570769	Female	26	80000	0
## 12	15606274	Female	26	52000	0
## 13	15746139	Male	20	86000	0
## 14	15704987	Male	32	18000	0
## 15	15628972	Male	18	82000	0
## 16	15697686	Male	29	80000	0
## 17	15733883	Male	47	25000	1
## 18	15617482	Male	45	26000	1
## 19	15704583	Male	46	28000	1
## 20	15621083	Female	48	29000	1
## 21	15649487	Male	45	22000	1
## 22	15736760	Female	47	49000	1
## 23	15714658	Male	48	41000	1
## 24	15599081	Female	45	22000	1
## 25	15705113	Male	46	23000	1
## 26	15631159	Male	47	20000	1
## 27	15792818	Male	49	28000	1
## 28	15633531	Female	47	30000	1
## 29	15744529	Male	29	43000	0
## 30	15669656	Male	31	18000	0

##	31	15581198	Male	31	74000	0
##	32	15729054	Female	27	137000	1
##	33	15573452	Female	21	16000	0
##	34	15776733	Female	28	44000	0
##	35	15724858	Male	27	90000	0
##	36	15713144	Male	35	27000	0
##	37	15690188	Female	33	28000	0
##	38	15689425	Male	30	49000	0
##	39	15671766	Female	26	72000	0
##	40	15782806	Female	27	31000	0
##	41	15764419	Female	27	17000	0
##	42	15591915	Female	33	51000	0
##	43	15772798	Male	35	108000	0
##	44	15792008	Male	30	15000	0
##	45	15715541	Female	28	84000	0
##	46	15639277	Male	23	20000	0
##	47	15798850	Male	25	79000	0
##	48	15776348	Female	27	54000	0
##	49	15727696	Male	30	135000	1
##	50	15793813	Female	31	89000	0
##	51	15694395	Female	24	32000	0
##	52	15764195	Female	18	44000	0
##	53	15744919	Female	29	83000	0
##	54	15671655	Female	35	23000	0
##	55	15654901	Female	27	58000	0
##	56	15649136	Female	24	55000	0
##	57	15775562	Female	23	48000	0
##	58	15807481	Male	28	79000	0
##	59	15642885	Male	22	18000	0
##	60	15789109	Female	32	117000	0
##	61	15814004	Male	27	20000	0
##	62	15673619	Male	25	87000	0

##	63	15595135	Female	23	66000	0
##	64	15583681	Male	32	120000	1
##	65	15605000	Female	59	83000	0
##	66	15718071	Male	24	58000	0
##	67	15679760	Male	24	19000	0
##	68	15654574	Female	23	82000	0
##	69	15577178	Female	22	63000	0
##	70	15595324	Female	31	68000	0
##	71	15756932	Male	25	80000	0
##	72	15726358	Female	24	27000	0
##	73	15595228	Female	20	23000	0
##	74	15782530	Female	33	113000	0
##	75	15592877	Male	32	18000	0
##	76	15651983	Male	34	112000	1
##	77	15746737	Male	18	52000	0
##	78	15774179	Female	22	27000	0
##	79	15667265	Female	28	87000	0
##	80	15655123	Female	26	17000	0
##	81	15595917	Male	30	80000	0
##	82	15668385	Male	39	42000	0
##	83	15709476	Male	20	49000	0
##	84	15711218	Male	35	88000	0
##	85	15798659	Female	30	62000	0
##	86	15663939	Female	31	118000	1
##	87	15694946	Male	24	55000	0
##	88	15631912	Female	28	85000	0
##	89	15768816	Male	26	81000	0
##	90	15682268	Male	35	50000	0
##	91	15684801	Male	22	81000	0
##	92	15636428	Female	30	116000	0
##	93	15809823	Male	26	15000	0
##	94	15699284	Female	29	28000	0

##	95	15786993	Female	29	83000	0
##	96	15709441	Female	35	44000	0
##	97	15710257	Female	35	25000	0
##	98	15582492	Male	28	123000	1
##	99	15575694	Male	35	73000	0
##	100	15756820	Female	28	37000	0
##	101	15766289	Male	27	88000	0
##	102	15593014	Male	28	59000	0
##	103	15584545	Female	32	86000	0
##	104	15675949	Female	33	149000	1
##	105	15672091	Female	19	21000	0
##	106	15801658	Male	21	72000	0
##	107	15706185	Female	26	35000	0
##	108	15789863	Male	27	89000	0
##	109	15720943	Male	26	86000	0
##	110	15697997	Female	38	80000	0
##	111	15665416	Female	39	71000	0
##	112	15660200	Female	37	71000	0
##	113	15619653	Male	38	61000	0
##	114	15773447	Male	37	55000	0
##	115	15739160	Male	42	80000	0
##	116	15689237	Male	40	57000	0
##	117	15679297	Male	35	75000	0
##	118	15591433	Male	36	52000	0
##	119	15642725	Male	40	59000	0
##	120	15701962	Male	41	59000	0
##	121	15811613	Female	36	75000	0
##	122	15741049	Male	37	72000	0
##	123	15724423	Female	40	75000	0
##	124	15574305	Male	35	53000	0
##	125	15678168	Female	41	51000	0
##	126	15697020	Female	39	61000	0

##	127	15610801	Male	42	65000	0
##	128	15745232	Male	26	32000	0
##	129	15722758	Male	30	17000	0
##	130	15792102	Female	26	84000	0
##	131	15675185	Male	31	58000	0
##	132	15801247	Male	33	31000	0
##	133	15725660	Male	30	87000	0
##	134	15638963	Female	21	68000	0
##	135	15800061	Female	28	55000	0
##	136	15578006	Male	23	63000	0
##	137	15668504	Female	20	82000	0
##	138	15687491	Male	30	107000	1
##	139	15610403	Female	28	59000	0
##	140	15741094	Male	19	25000	0
##	141	15807909	Male	19	85000	0
##	142	15666141	Female	18	68000	0
##	143	15617134	Male	35	59000	0
##	144	15783029	Male	30	89000	0
##	145	15622833	Female	34	25000	0
##	146	15746422	Female	24	89000	0
##	147	15750839	Female	27	96000	1
##	148	15749130	Female	41	30000	0
##	149	15779862	Male	29	61000	0
##	150	15767871	Male	20	74000	0
##	151	15679651	Female	26	15000	0
##	152	15576219	Male	41	45000	0
##	153	15699247	Male	31	76000	0
##	154	15619087	Female	36	50000	0
##	155	15605327	Male	40	47000	0
##	156	15610140	Female	31	15000	0
##	157	15791174	Male	46	59000	0
##	158	15602373	Male	29	75000	0

##	159	15762605	Male	26	30000	0
##	160	15598840	Female	32	135000	1
##	161	15744279	Male	32	100000	1
##	162	15670619	Male	25	90000	0
##	163	15599533	Female	37	33000	0
##	164	15757837	Male	35	38000	0
##	165	15697574	Female	33	69000	0
##	166	15578738	Female	18	86000	0
##	167	15762228	Female	22	55000	0
##	168	15614827	Female	35	71000	0
##	169	15789815	Male	29	148000	1
##	170	15579781	Female	29	47000	0
##	171	15587013	Male	21	88000	0
##	172	15570932	Male	34	115000	0
##	173	15794661	Female	26	118000	0
##	174	15581654	Female	34	43000	0
##	175	15644296	Female	34	72000	0
##	176	15614420	Female	23	28000	0
##	177	15609653	Female	35	47000	0
##	178	15594577	Male	25	22000	0
##	179	15584114	Male	24	23000	0
##	180	15673367	Female	31	34000	0
##	181	15685576	Male	26	16000	0
##	182	15774727	Female	31	71000	0
##	183	15694288	Female	32	117000	1
##	184	15603319	Male	33	43000	0
##	185	15759066	Female	33	60000	0
##	186	15814816	Male	31	66000	0
##	187	15724402	Female	20	82000	0
##	188	15571059	Female	33	41000	0
##	189	15674206	Male	35	72000	0
##	190	15715160	Male	28	32000	0

##	191	15730448	Male	24	84000	0
##	192	15662067	Female	19	26000	0
##	193	15779581	Male	29	43000	0
##	194	15662901	Male	19	70000	0
##	195	15689751	Male	28	89000	0
##	196	15667742	Male	34	43000	0
##	197	15738448	Female	30	79000	0
##	198	15680243	Female	20	36000	0
##	199	15745083	Male	26	80000	0
##	200	15708228	Male	35	22000	0
##	201	15628523	Male	35	39000	0
##	202	15708196	Male	49	74000	0
##	203	15735549	Female	39	134000	1
##	204	15809347	Female	41	71000	0
##	205	15660866	Female	58	101000	1
##	206	15766609	Female	47	47000	0
##	207	15654230	Female	55	130000	1
##	208	15794566	Female	52	114000	0
##	209	15800890	Female	40	142000	1
##	210	15697424	Female	46	22000	0
##	211	15724536	Female	48	96000	1
##	212	15735878	Male	52	150000	1
##	213	15707596	Female	59	42000	0
##	214	15657163	Male	35	58000	0
##	215	15622478	Male	47	43000	0
##	216	15779529	Female	60	108000	1
##	217	15636023	Male	49	65000	0
##	218	15582066	Male	40	78000	0
##	219	15666675	Female	46	96000	0
##	220	15732987	Male	59	143000	1
##	221	15789432	Female	41	80000	0
##	222	15663161	Male	35	91000	1

##	223	15694879	Male	37	144000	1
##	224	15593715	Male	60	102000	1
##	225	15575002	Female	35	60000	0
##	226	15622171	Male	37	53000	0
##	227	15795224	Female	36	126000	1
##	228	15685346	Male	56	133000	1
##	229	15691808	Female	40	72000	0
##	230	15721007	Female	42	80000	1
##	231	15794253	Female	35	147000	1
##	232	15694453	Male	39	42000	0
##	233	15813113	Male	40	107000	1
##	234	15614187	Male	49	86000	1
##	235	15619407	Female	38	112000	0
##	236	15646227	Male	46	79000	1
##	237	15660541	Male	40	57000	0
##	238	15753874	Female	37	80000	0
##	239	15617877	Female	46	82000	0
##	240	15772073	Female	53	143000	1
##	241	15701537	Male	42	149000	1
##	242	15736228	Male	38	59000	0
##	243	15780572	Female	50	88000	1
##	244	15769596	Female	56	104000	1
##	245	15586996	Female	41	72000	0
##	246	15722061	Female	51	146000	1
##	247	15638003	Female	35	50000	0
##	248	15775590	Female	57	122000	1
##	249	15730688	Male	41	52000	0
##	250	15753102	Female	35	97000	1
##	251	15810075	Female	44	39000	0
##	252	15723373	Male	37	52000	0
##	253	15795298	Female	48	134000	1
##	254	15584320	Female	37	146000	1

##	255	15724161	Female	50	44000	0
##	256	15750056	Female	52	90000	1
##	257	15609637	Female	41	72000	0
##	258	15794493	Male	40	57000	0
##	259	15569641	Female	58	95000	1
##	260	15815236	Female	45	131000	1
##	261	15811177	Female	35	77000	0
##	262	15680587	Male	36	144000	1
##	263	15672821	Female	55	125000	1
##	264	15767681	Female	35	72000	0
##	265	15600379	Male	48	90000	1
##	266	15801336	Female	42	108000	1
##	267	15721592	Male	40	75000	0
##	268	15581282	Male	37	74000	0
##	269	15746203	Female	47	144000	1
##	270	15583137	Male	40	61000	0
##	271	15680752	Female	43	133000	0
##	272	15688172	Female	59	76000	1
##	273	15791373	Male	60	42000	1
##	274	15589449	Male	39	106000	1
##	275	15692819	Female	57	26000	1
##	276	15727467	Male	57	74000	1
##	277	15734312	Male	38	71000	0
##	278	15764604	Male	49	88000	1
##	279	15613014	Female	52	38000	1
##	280	15759684	Female	50	36000	1
##	281	15609669	Female	59	88000	1
##	282	15685536	Male	35	61000	0
##	283	15750447	Male	37	70000	1
##	284	15663249	Female	52	21000	1
##	285	15638646	Male	48	141000	0
##	286	15734161	Female	37	93000	1

##	287	15631070	Female	37	62000	0
##	288	15761950	Female	48	138000	1
##	289	15649668	Male	41	79000	0
##	290	15713912	Female	37	78000	1
##	291	15586757	Male	39	134000	1
##	292	15596522	Male	49	89000	1
##	293	15625395	Male	55	39000	1
##	294	15760570	Male	37	77000	0
##	295	15566689	Female	35	57000	0
##	296	15725794	Female	36	63000	0
##	297	15673539	Male	42	73000	1
##	298	15705298	Female	43	112000	1
##	299	15675791	Male	45	79000	0
##	300	15747043	Male	46	117000	1
##	301	15736397	Female	58	38000	1
##	302	15678201	Male	48	74000	1
##	303	15720745	Female	37	137000	1
##	304	15637593	Male	37	79000	1
##	305	15598070	Female	40	60000	0
##	306	15787550	Male	42	54000	0
##	307	15603942	Female	51	134000	0
##	308	15733973	Female	47	113000	1
##	309	15596761	Male	36	125000	1
##	310	15652400	Female	38	50000	0
##	311	15717893	Female	42	70000	0
##	312	15622585	Male	39	96000	1
##	313	15733964	Female	38	50000	0
##	314	15753861	Female	49	141000	1
##	315	15747097	Female	39	79000	0
##	316	15594762	Female	39	75000	1
##	317	15667417	Female	54	104000	1
##	318	15684861	Male	35	55000	0

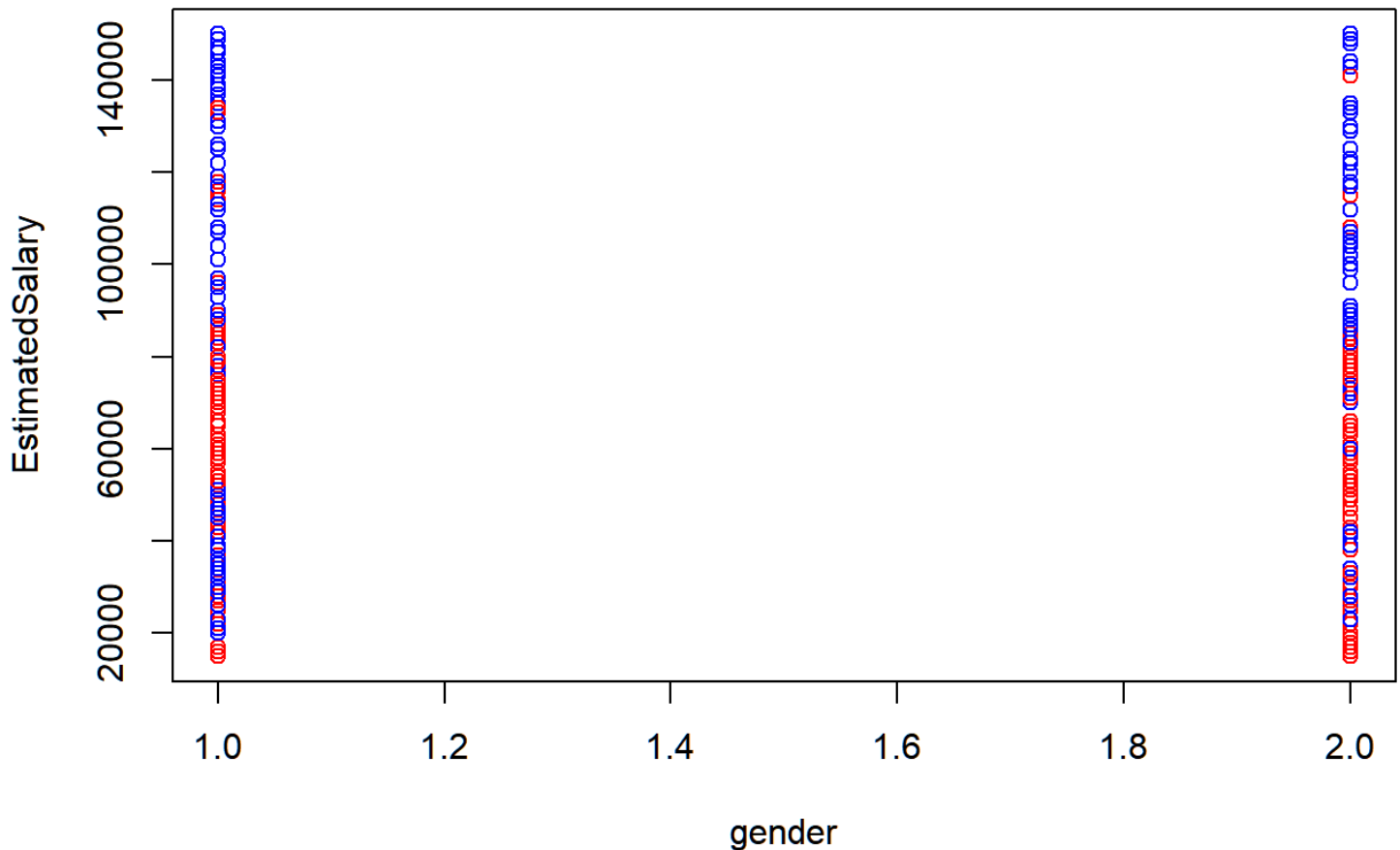
##	319	15742204	Male	45	32000	1
##	320	15623502	Male	36	60000	0
##	321	15774872	Female	52	138000	1
##	322	15611191	Female	53	82000	1
##	323	15674331	Male	41	52000	0
##	324	15619465	Female	48	30000	1
##	325	15575247	Female	48	131000	1
##	326	15695679	Female	41	60000	0
##	327	15713463	Male	41	72000	0
##	328	15785170	Female	42	75000	0
##	329	15796351	Male	36	118000	1
##	330	15639576	Female	47	107000	1
##	331	15693264	Male	38	51000	0
##	332	15589715	Female	48	119000	1
##	333	15769902	Male	42	65000	0
##	334	15587177	Male	40	65000	0
##	335	15814553	Male	57	60000	1
##	336	15601550	Female	36	54000	0
##	337	15664907	Male	58	144000	1
##	338	15612465	Male	35	79000	0
##	339	15810800	Female	38	55000	0
##	340	15665760	Male	39	122000	1
##	341	15588080	Female	53	104000	1
##	342	15776844	Male	35	75000	0
##	343	15717560	Female	38	65000	0
##	344	15629739	Female	47	51000	1
##	345	15729908	Male	47	105000	1
##	346	15716781	Female	41	63000	0
##	347	15646936	Male	53	72000	1
##	348	15768151	Female	54	108000	1
##	349	15579212	Male	39	77000	0
##	350	15721835	Male	38	61000	0

##	351	15800515	Female	38	113000	1
##	352	15591279	Male	37	75000	0
##	353	15587419	Female	42	90000	1
##	354	15750335	Female	37	57000	0
##	355	15699619	Male	36	99000	1
##	356	15606472	Male	60	34000	1
##	357	15778368	Male	54	70000	1
##	358	15671387	Female	41	72000	0
##	359	15573926	Male	40	71000	1
##	360	15709183	Male	42	54000	0
##	361	15577514	Male	43	129000	1
##	362	15778830	Female	53	34000	1
##	363	15768072	Female	47	50000	1
##	364	15768293	Female	42	79000	0
##	365	15654456	Male	42	104000	1
##	366	15807525	Female	59	29000	1
##	367	15574372	Female	58	47000	1
##	368	15671249	Male	46	88000	1
##	369	15779744	Male	38	71000	0
##	370	15624755	Female	54	26000	1
##	371	15611430	Female	60	46000	1
##	372	15774744	Male	60	83000	1
##	373	15629885	Female	39	73000	0
##	374	15708791	Male	59	130000	1
##	375	15793890	Female	37	80000	0
##	376	15646091	Female	46	32000	1
##	377	15596984	Female	46	74000	0
##	378	15800215	Female	42	53000	0
##	379	15577806	Male	41	87000	1
##	380	15749381	Female	58	23000	1
##	381	15683758	Male	42	64000	0
##	382	15670615	Male	48	33000	1

##	383	15715622	Female	44	139000	1
##	384	15707634	Male	49	28000	1
##	385	15806901	Female	57	33000	1
##	386	15775335	Male	56	60000	1
##	387	15724150	Female	49	39000	1
##	388	15627220	Male	39	71000	0
##	389	15672330	Male	47	34000	1
##	390	15668521	Female	48	35000	1
##	391	15807837	Male	48	33000	1
##	392	15592570	Male	47	23000	1
##	393	15748589	Female	45	45000	1
##	394	15635893	Male	60	42000	1
##	395	15757632	Female	39	59000	0
##	396	15691863	Female	46	41000	1
##	397	15706071	Male	51	23000	1
##	398	15654296	Female	50	20000	1
##	399	15755018	Male	36	33000	0
##	400	15594041	Female	49	36000	1

b) 根据性别和收入变量，对有购买欲的和无购买欲的用户使用不同的颜色画出对应的散点图，观察分类情况。

```
df=data.frame(gender=as.numeric( ads$Gender), ads[4])
df$Colour[ads$Purchased==0]="red"
df$Colour[ads$Purchased==1]="blue"
plot(df[1:2], col=df$Colour, type="p")
```



c) 使用线性支持向量机进行分类并尝试评价模型准确率

```
y= ads$Purchased  
dat = df[1:2]  
svmfit = svm( y~ . , data = dat, kernel = "linear", cost = 10,  
  scale = FALSE)  
plot(svmfit)
```

d) 尝试使用非线性支持向量机完成上述实验


```
y= ads$Purchased
dat = df[1:2]
radfit = svm(factor(y) ~ ., data = dat, scale = FALSE, kernel =
"radial", cost = 5)
# plot(radfit)
```

2. 实验二

a) 将 Position_Salaries 数据导入编程语言

```
ps <- read.csv(file = '../data/Position_Salaries.csv')
ps
```

##		Position Level	Salary
## 1	Business Analyst	1	45000
## 2	Junior Consultant	2	50000
## 3	Senior Consultant	3	60000
## 4	Manager	4	80000
## 5	Country Manager	5	110000
## 6	Region Manager	6	150000
## 7	Partner	7	200000
## 8	Senior Partner	8	300000
## 9	C-level	9	500000
## 10	CEO	10	1000000

b) 根据线性回归与职业等级作线性回归，并求得线性回归的 R-Squared

```
x=unlist(ps[2])
y=unlist(ps[3])
rel=lm(y ~ x)
summary(rel)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -170818 -129720  -40379   65856  386545
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -195333    124790   -1.565   0.15615
## x              80879     20112    4.021   0.00383 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 182700 on 8 degrees of freedom
## Multiple R-squared:  0.669,    Adjusted R-squared:  0.6277
## F-statistic: 16.17 on 1 and 8 DF,  p-value: 0.003833
```

c) 预测 level 为 4.5 和 8.5 的员工薪水

```
a <- data.frame(x = 4.5)
result1 <- predict(rel, a)
```

```
a <- data.frame(x = 8.5)
result2 <- predict(rel, a)
print(c(result1, result2))
```

```
##           1           1
## 168621.2 492136.4
```

3. 实验三

a) 将实验二的线性回归方程利用在 Social_Network_Ads 数据数据中，得到每个用户的职业等级

```
inverse.predict(rel, ads$EstimatedSalary)
```

```
## $Prediction
## [1] 3.277446
##
## $`Standard Error`
## [1] 0.9101292
##
## $Confidence
## [1] 2.098762
##
## $`Confidence Limits`
## [1] 1.178684 5.376207
```

b) 尝试使用支持向量机对三维的数据进行分类。

```
svmfit = svm( ads$Purchased ~ . , data = ads[2:4], kernel = "linear", cost = 10, scale = FALSE)
```