In [1]:
```
pip install seaborn
```

```
Requirement already satisfied: seaborn in c:\users\lenovo\appdata\local\progr
ams\python\python311\lib\site-packages (0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\users\lenovo\appdat
a\local\programs\python\python311\lib\site-packages (from seaborn) (1.24.3)
Requirement already satisfied: pandas>=0.25 in c:\users\lenovo\appdata\local
\programs\python\python311\lib\site-packages (from seaborn) (2.0.1)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\users\lenovo\app
data\local\programs\python\python311\lib\site-packages (from seaborn) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\lenovo\appdata\lo
cal\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1
->seaborn) (1.0.7)
Requirement already satisfied: cycler>=0.10 in c:\users\lenovo\appdata\local
\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1->s
eaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\lenovo\appdata\l
ocal\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.
1->seaborn) (4.39.4)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\lenovo\appdata\l
ocal\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.
1->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\lenovo\appdata\loc
al\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1-
>seaborn) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\lenovo\appdata\local
\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1->s
eaborn) (9.5.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\lenovo\appdata\lo
cal\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1
->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\lenovo\appdat
a\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>
=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\lenovo\appdata\local
\programs\python\python311\lib\site-packages (from pandas>=0.25->seaborn) (20
23.3)
Requirement already satisfied: tzdata>=2022.1 in c:\users\lenovo\appdata\loca
l\programs\python\python311\lib\site-packages (from pandas>=0.25->seaborn) (2
023.3)
Requirement already satisfied: six>=1.5 in c:\users\lenovo\appdata\local\prog
rams\python\python311\lib\site-packages (from python-dateutil>=2.7->matplotli
b!=3.6.1,>=3.1->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
pip install --upgrade pip
```

```
Requirement already satisfied: pip in c:\users\lenovo\appdata\local\programs
\python\python311\lib\site-packages (23.1.2)
Note: you may need to restart the kernel to use updated packages.
```

In [3]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn import preprocessing,svm
from sklearn.linear_model import LinearRegression
```

In [4]:
```
df=pd.read_csv(r"C:\Users\Lenovo\OneDrive\Desktop\Data Sets\bottle.csv")
df
```

```
C:\Users\Lenovo\AppData\Local\Temp\ipykernel_4356\1599601436.py:1: DtypeWarni
ng: Columns (47,73) have mixed types. Specify dtype option on import or set l
ow_memory=False.
  df=pd.read_csv(r"C:\Users\Lenovo\OneDrive\Desktop\Data Sets\bottle.csv")
```

Out[4]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sa
0	1	1	054.0 056.0	19-4903CR-HY-060-0930-05400560-0000A-3	0	10.500	33.4400	NaN	25.64900	Na
1	1	2	054.0 056.0	19-4903CR-HY-060-0930-05400560-0008A-3	8	10.460	33.4400	NaN	25.65600	Na
2	1	3	054.0 056.0	19-4903CR-HY-060-0930-05400560-0010A-7	10	10.460	33.4370	NaN	25.65400	Na
3	1	4	054.0 056.0	19-4903CR-HY-060-0930-05400560-0019A-3	19	10.450	33.4200	NaN	25.64300	Na
4	1	5	054.0 056.0	19-4903CR-HY-060-0930-05400560-0020A-7	20	10.450	33.4210	NaN	25.64300	Na
...	...	...	...	...	...	...	...	...	...	.
864858	34404	864859	093.4 026.4	20-1611SR-MX-310-2239-09340264-0000A-7	0	18.744	33.4083	5.805	23.87055	108.7
864859	34404	864860	093.4 026.4	20-1611SR-MX-310-2239-09340264-0002A-3	2	18.744	33.4083	5.805	23.87072	108.7
864860	34404	864861	093.4 026.4	20-1611SR-MX-310-2239-09340264-0005A-3	5	18.692	33.4150	5.796	23.88911	108.4
864861	34404	864862	093.4 026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	18.161	33.4062	5.816	24.01426	107.7

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sa
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774	24.15297	105.6

864863 rows × 74 columns

In [5]:
```
df= df[['Salnty','T_degC']]
df.columns=['Sal','Temp']
```

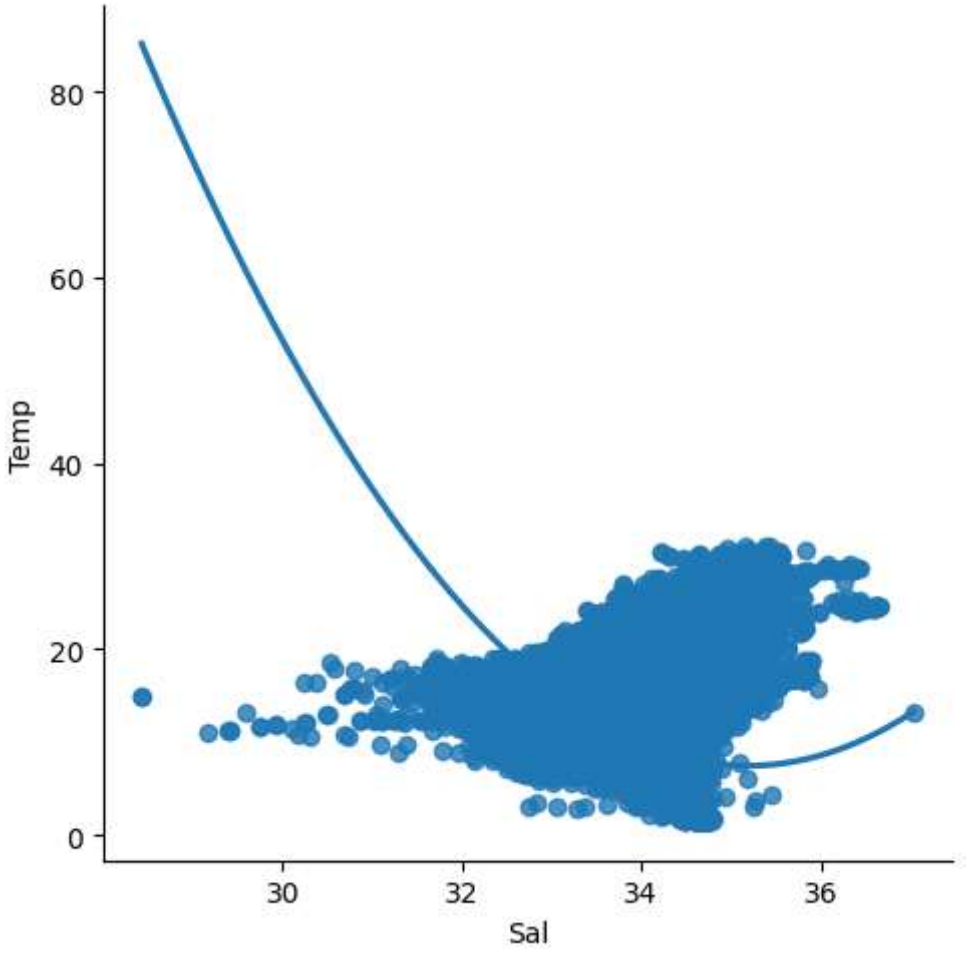
In [6]:
```
df.head(10)
```

Out[6]:

	Sal	Temp
0	33.440	10.50
1	33.440	10.46
2	33.437	10.46
3	33.420	10.45
4	33.421	10.45
5	33.431	10.45
6	33.440	10.45
7	33.424	10.24
8	33.420	10.06
9	33.494	9.86

In [7]:
```
sns.lmplot(x="Sal",y="Temp",data=df,order=2,ci=None)
```

Out[7]:
```
<seaborn.axisgrid.FacetGrid at 0x29de22c0310>
```



In [8]:
```
df.describe()
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 864863 entries, 0 to 864862
Data columns (total 2 columns):
 #   Column  Non-Null Count   Dtype  
---  ------  --------------   -----  
 0   Sal     817509 non-null  float64
 1   Temp    853900 non-null  float64
dtypes: float64(2)
memory usage: 13.2 MB
```

In [9]:
```
df.fillna(method='ffill',inplace=True)
```

```
C:\Users\Lenovo\AppData\Local\Temp\ipykernel_4356\4116506308.py:1: SettingWit
hCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  df.fillna(method='ffill',inplace=True)
```

In [10]:
```
regr=LinearRegression()
```

In [12]:
```
x=np.array(df['Sal']).reshape(-1,1)
y=np.array(df['Temp']).reshape(-1,1)
df.dropna(inplace=True)
```

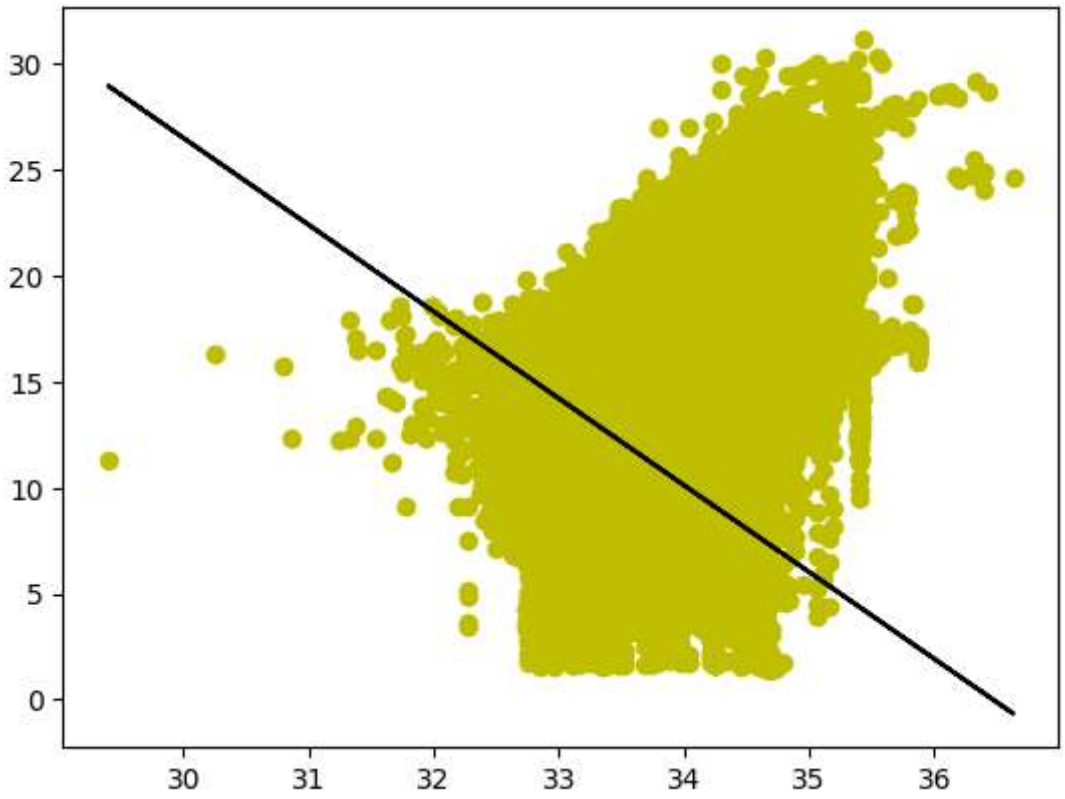
```
C:\Users\Lenovo\AppData\Local\Temp\ipykernel_4356\1516682253.py:3: SettingWit
hCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  df.dropna(inplace=True)
```

In [19]:
```
X_train,X_test,Y_train,Y_test=train_test_split(x,y,test_size=0.25)
regr.fit(X_train,Y_train)
regr.fit(X_train,Y_train)
```

Out[19]:
```
LinearRegression()
```

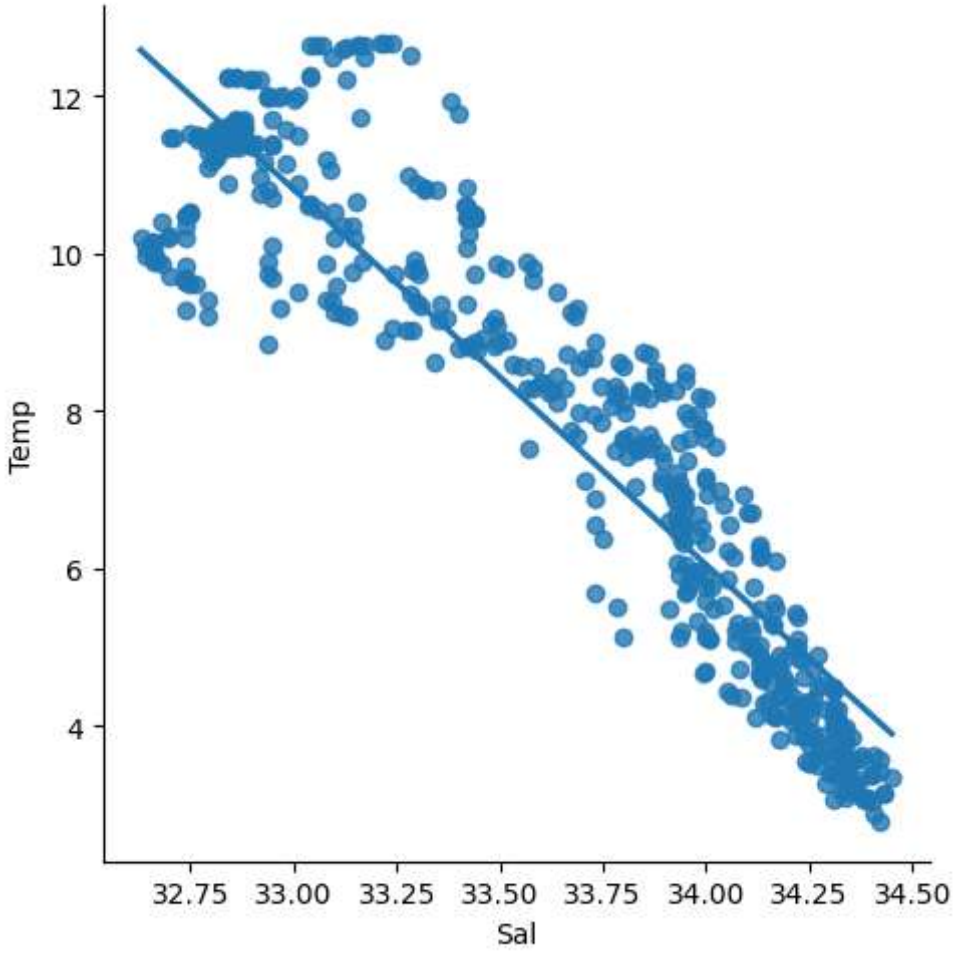
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [21]:
```
y_pred=regr.predict(X_test)
plt.scatter(X_test,Y_test,color='y')
plt.plot(X_test,y_pred,color='k')
plt.show()
```



In [25]:
```
df500=df[:][:500]
sns.lmplot(x="Sal",y="Temp",data=df500,order=1,ci=None)
```

Out[25]: <seaborn.axisgrid.FacetGrid at 0x29d84c4c750>



In [27]:
```
df500.fillna(method='ffill',inplace=True)
x=np.array(df500['Sal']).reshape(-1,1)
y=np.array(df500['Temp']).reshape(-1,1)
df500.dropna(inplace=True)
X_train,X_test,Y_train,Y_test=train_test_split(x,y,test_size=0.25)
```

In [28]:
```
regr.fit(X_train,Y_train)
```

Out[28]:
```
LinearRegression()
```

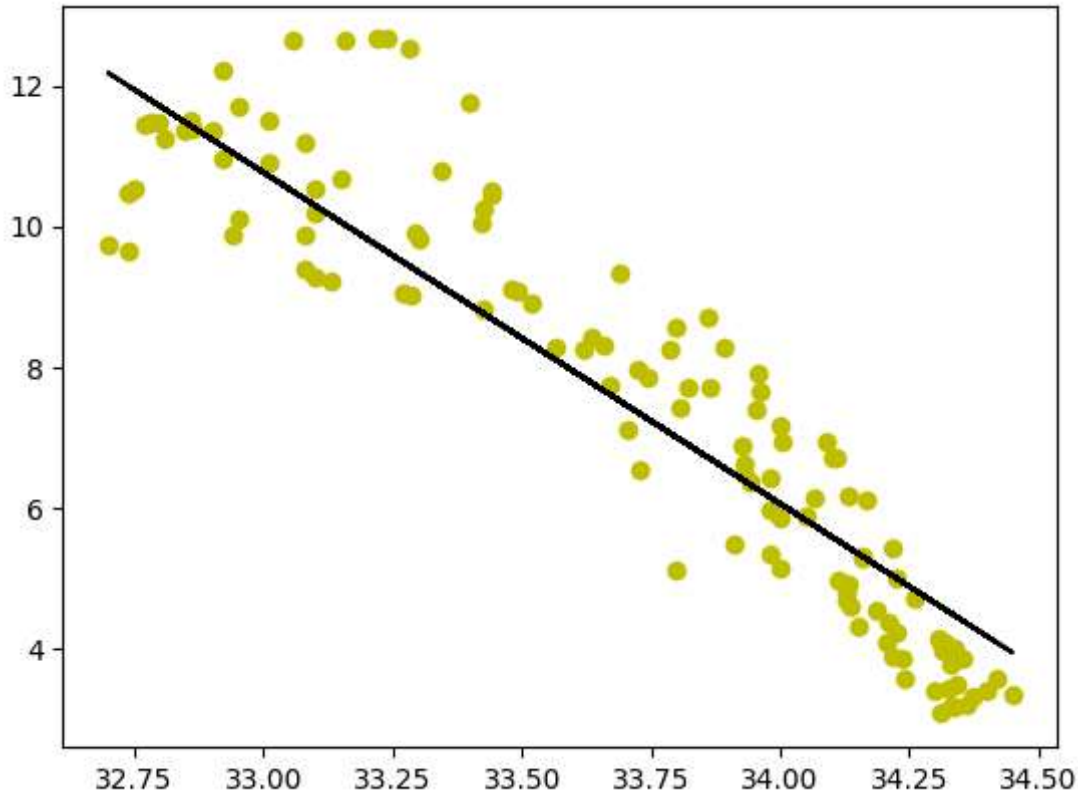
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [29]:
```
print("Regression:",regr.score(X_test,Y_test))
```

```
Regression: 0.8520118787637433
```

In [30]:
```
y_pred=regr.predict(X_test)
```

In [33]:
```
plt.scatter(X_test,Y_test,color='y')
plt.plot(X_test,y_pred,color='k')
plt.show()
```



In [35]:
```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import  r2_score
model=LinearRegression()
model.fit(X_train,Y_train)
```

Out[35]:
```
LinearRegression()
```
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [36]:
```
y_pred=model.predict(X_test)
```

In [37]:
```
r2=r2_score(Y_test,y_pred)
```

In [38]:
```
print("R2 Score:",r2)
```

```
R2 Score: 0.8520118787637433
```